

Hybrid AI-physics screening prioritizes candidate dual MDM2/MDM4 inhibitors from a drug-like library for dedifferentiated liposarcoma

Solomon Sackitey

Southern Illinois University Edwardsville

Prince Danan Biniyam

2016pbiniyam@uhas.edu.gh

University of Health and Allied Sciences

Charles Lamptey Mireku

KNUST

Alberta Hammah

University of Health and Allied Sciences

Jesse Ayim Buabeng

University of Health and Allied Sciences

Lawrence Kwame Nkrumah

KNUST

Wise Mawuli Asare

University of Ghana

Josephine Frimpomaa Oppong

Central University

Emmanuel Gyamfi Obeng

University of Health and Allied Sciences

Kwabena Adu-Adjei

University of Health and Allied Sciences

Clifford Tignangnen Penajah

University of Health and Allied Sciences

Cedric Dzidzor Kodjo Amengor

University of Health and Allied Sciences

Research Article

Keywords: Dedifferentiated liposarcoma, MDM2/MDM4 inhibitors, ensemble docking, machine learning QSAR, molecular dynamics, MM/GBSA, network pharmacology, virtual screening

Posted Date: September 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10832690/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Hybrid AI-physics screening prioritizes candidate dual MDM2/MDM4 inhibitors from a drug-like library for dedifferentiated liposarcoma.

Solomon Sackitey¹, Prince Danan Biniyam^{2*}, Charles Lamptey Mireku³, Alberta Hammah², Jesse Ayim Buabeng², Lawrence Kwame Nkrumah³, Wise Mawuli Asare⁴, Josephine Frimpomaa Oppong⁵, Emmanuel Gyamfi Obeng², Kwabena Adu-Adjei², Clifford Tignangnen Penajah², Cedric Dzidzor Kodjo Amengor²

Author Affiliations:

¹Department of Chemistry, Southern Illinois University Edwardsville, USA.

²Department of Pharmaceutical Chemistry, School of Pharmacy, University of Health and Allied Sciences, Ho, Ghana.

³Department of Pharmaceutical chemistry, Faculty of Pharmacy and pharmaceutical Science, KNUST, Kumasi, Ghana

⁴Department of Biomedical Engineering, school of Engineering sciences, University of Ghana, Accra, Ghana.

⁵Department of Pharmaceutical Sciences, School of Pharmacy, Central University, Accra, Ghana

Corresponding Author

Prince Danan Biniyam

Department of Pharmaceutical Chemistry, Drug Discovery Unit, School of Pharmacy, University of Health and Allied Sciences, Ho, Ghana.

Email: 2016pbiniyam@uhas.edu.gh

Contact: +233 59 429 7607

Abstract

Dedifferentiated liposarcoma (DDLPS) is driven by MDM2 and CDK4 co-amplification with wild-type TP53, establishing the MDM2-p53 axis as a therapeutic target. Clinical MDM2-selective inhibitors show limited efficacy, partly due to compensatory MDM4 upregulation. We present an integrated pipeline combining ensemble machine-learning QSAR, ensemble docking across 23 MDM2 and 24 MDM4 crystal structures, molecular dynamics simulations, MM/GBSA free-energy calculations, network pharmacology, and ADMET profiling to prioritize candidate dual MDM2/MDM4 inhibitors. Screening a drug-like subset of 9.4 million ZINC20 compounds yielded two structurally novel candidate compounds, ZINC9412756 and ZINC20724474, with predicted MDM2 IC₅₀ < 10 nM. ECFP4-based Tanimoto similarity analysis against known MDM2 (n = 4,154) and MDM4 (n = 155) inhibitors indicated that both compounds are structurally distinct from known inhibitors, with maximum similarities of 0.3827 and 0.3488 for MDM2, and 0.2500 and 0.2315 for MDM4, respectively, all well below the 0.5 threshold for structural relatedness. Ensemble docking predicted mean binding affinities of -9.10 to -9.27 kcal/mol for both targets, comparable to or exceeding those of co-crystallised reference inhibitors. Molecular dynamics (500 ns) suggested stable binding, driven predominantly by hydrophobic packing; van der Waals interactions contributed 70-90% of the gas-phase stabilization in MM/GBSA, with persistent hydrogen bonds providing secondary anchoring. MM/GBSA yielded binding free energies of -20.73 ± 4.02 to -26.60 ± 2.76 kcal/mol. Network pharmacology predicted engagement of a DDLPS-relevant module (MDM2, MDM4, CDK4, EGFR, AURKA) with significant enrichment of p53 signaling (FDR = 1.42×10^{-5}). ADMET predictions indicated favorable oral bioavailability and synthetic accessibility, with several toxicity alerts that represent tractable optimization targets. This integrated AI-physics approach efficiently prioritizes structurally novel candidate dual MDM2/MDM4 inhibitors with polypharmacological potential for DDLPS, though experimental validation is required to confirm binding affinities and functional activity.

Keywords: Dedifferentiated liposarcoma; MDM2/MDM4 inhibitors; ensemble docking; machine learning QSAR; molecular dynamics; MM/GBSA; network pharmacology; virtual screening.

Highlights

- Integrated ML-QSAR, ensemble docking, MD/MMGBSA, network pharmacology and ADMET pipeline for dual MDM2/MDM4 inhibitor prioritization.
- Random Forest QSAR achieved test $R^2 = 0.76$, RMSE = 0.64, CCC = 0.87, with y-randomization indicating genuine structure-activity relationships.
- Candidate compounds ZINC9412756 and ZINC20724474 (predicted MDM2 IC₅₀ < 10 nM) showed docking scores of -9.10 ± 0.30 to -9.27 ± 0.30 kcal/mol for both targets, exceeding those of co-crystallised reference inhibitors. ECFP4-based structural similarity analysis against >4,000 known MDM2/MDM4 inhibitors indicated their structural novelty (max Tanimoto = 0.38 for MDM2, = 0.25 for MDM4).
- MD simulations and MM/GBSA suggested stable hydrophobic-dominated binding, with van der Waals interactions contributing 70-90% of the stabilization and MM/GBSA energies of -20.73 to -26.60 kcal/mol.
- Network pharmacology linked the candidate compounds to a DDLPS module (MDM2, MDM4, CDK4, EGFR, AURKA) enriched in p53 signaling (FDR = 1.42×10^{-5}), providing rationale for polypharmacological targeting.

1.0. Introduction

Dedifferentiated liposarcoma (DDLPS) is among the most prevalent soft-tissue sarcoma subtypes, accounting for approximately 10% of all sarcomas and marked by high rates of local recurrence and distant metastasis [1,2]. Surgical resection remains the standard of care, but negative margins are frequently difficult to achieve, and DDLPS demonstrates relative chemoresistance to anthracycline-based regimens, with radiographic responses observed in fewer than one-third of patients [3-5]. Consequently, outcomes remain poor, with median overall survival approximately 50% following the first local recurrence [6,7].

The molecular hallmark of DDLPS is co-amplification of *MDM2* and *CDK4* on chromosome 12q13-15, detected in over 90% of cases alongside wild-type *TP53* [8,9]. MDM2 functions as an E3 ubiquitin ligase that binds the transactivation domain of p53, inhibiting its transcriptional activity and targeting it for proteasomal degradation [10, 11] (**Fig.1a**). This establishes the MDM2-p53 interface as a central therapeutic vulnerability in DDLPS [12]. Genetic models demonstrate the essential role of MDM2 in p53 regulation; mice lacking *Mdm2* die during early embryogenesis from p53-dependent apoptosis, a phenotype rescued by *Trp53* deletion [13,14]. In human tumours, *MDM2* amplification and *TP53* mutation are largely mutually exclusive, indicating that MDM2 overexpression provides an alternative mechanism for p53 inactivation [9,15].

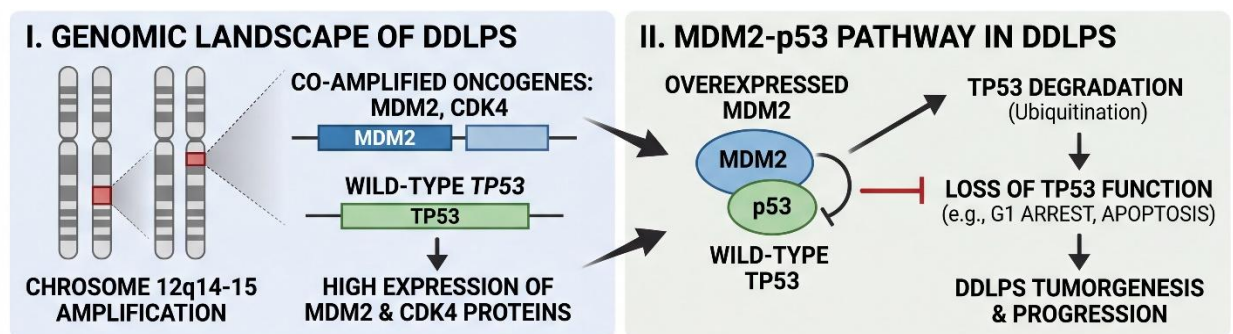


Fig. 1a. The MDM2-p53 axis in dedifferentiated liposarcoma

Pharmacological targeting of the MDM2-p53 axis began with nutlin-class compounds, which occupy the p53-binding cleft of MDM2 with nanomolar affinity [16,17]. Subsequent clinical candidates including RG7112, MI-77301, RG7388, AMG-232 and siremadlin have been evaluated in phase I/II trials [18-23]. However, selective MDM2 antagonists have shown limited single-agent activity in DDLPS, with low objective response rates and dose-limiting toxicities. In a phase I trial of RG7112, only one partial response was observed among 20 patients, with five developing progressive disease [24]. A phase I study of MI-77301 reported no objective responses, with stable disease as the best outcome [25]. Siremadlin yielded one partial response (8%) and nine cases of stable disease (75%) among 12 liposarcoma patients [23]. These data indicate that while MDM2 inhibition can stabilize disease in some patients, durable objective responses remain rare.

Resistance to MDM2-selective inhibitors arises through multiple mechanisms, most notably acquisition of *TP53* mutations that abolish p53 function [26,27]. Importantly, upregulation of the MDM2 homolog MDM4 (MDMX) can compensate for MDM2 inhibition by providing an alternative p53-binding partner [28-30] (**Fig.1b**). MDM4 binds p53 through a structurally analogous N-terminal domain but is not efficiently inhibited by nutlin-class compounds due to subtle binding cleft differences that reduce affinity [31]. MDM4 overexpression has been linked to poor responses to nutlin-3a in chronic lymphocytic leukaemia and solid tumour models [32,33], and MDM4 elevation in DDLPS has been associated with resistance to MDM2 inhibition [34]. Genetic suppression of MDM4 restores p53-dependent cell-cycle arrest and apoptosis [34], and the MDM2-MDM4 heterodimer is functionally distinct from the MDM2 homodimer, with MDM4 allosterically modulating MDM2 function and enhancing p53 degradation [29]. Preclinical studies suggest that simultaneous disruption of p53 binding to both MDM2 and MDM4 enhances p53 activation and overcomes MDM4-mediated resistance compared with selective MDM2 inhibition alone [35]. The dual inhibitor YL93 (CHEMBL5181559), with binding K_i values of 1.1 nM for MDM2 and 642 nM for MDM4, demonstrated superior activity in MDM4-overexpressing cancer cell lines compared to the MDM2-selective inhibitor RG7388 [36]. More recently, navtemadlin has shown promise in combination regimens [37]. These findings provide proof-of-principle that small-molecule dual antagonism of MDM2 and MDM4 is achievable, supporting continued exploration of this approach.

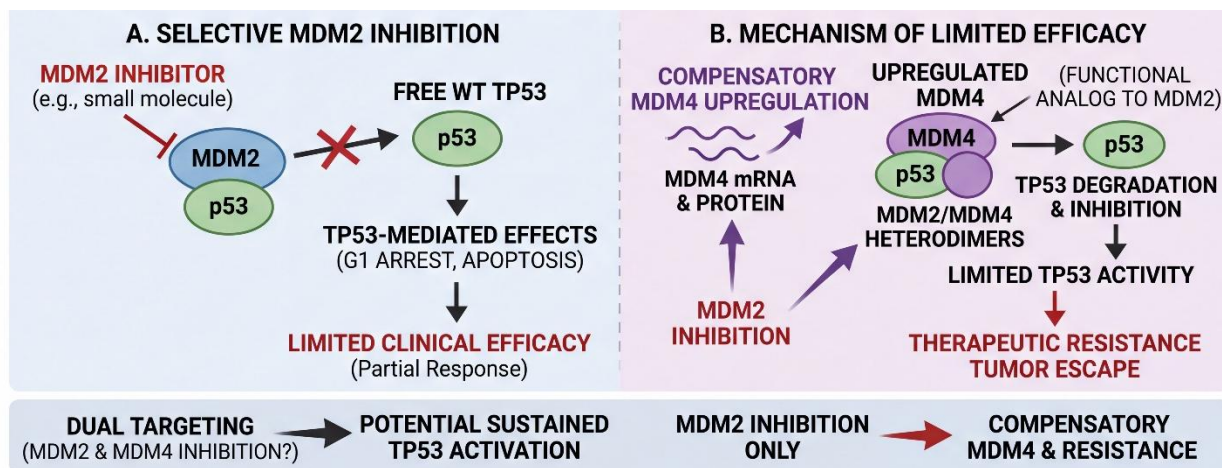


Fig. 1b. Therapeutic targeting of MDM2 and the compensatory resistance mechanism

Large, curated libraries of synthetically accessible compounds offer rich resources for discovering such dual inhibitors. The ZINC20 database contains over 230 million purchasable compounds and has been successfully employed for virtual screening against diverse targets [38-40]. Recent advances have enabled integration of artificial intelligence with physics-based simulations to efficiently navigate vast chemical spaces. Machine-learning QSAR models can predict compound potency with high accuracy, substantially reducing candidates requiring computationally expensive evaluation [41,42]. Ensemble docking across multiple receptor conformations accounts for protein flexibility, while molecular dynamics coupled with

MM/GBSA calculations yield robust binding free energy estimates [43]. The integration of these methodologies into a single automated pipeline represents the principal methodological advance of this study. This unified approach enables efficient navigation of ultra large chemical spaces (>9 million compounds), simultaneous assessment of binding to both MDM2 and MDM4 using structurally diverse receptor ensembles, evaluation of target engagement within a DDLPS-specific biological context, and prospective identification of ADMET liabilities prior to experimental investment [65-67].

In the present study, we deploy this integrated hybrid computational pipeline to screen a curated drug-like subset of the ZINC20 database, aiming to prioritize novel, synthetically accessible candidate dual MDM2/MDM4 inhibitors tailored to the molecular landscape of DDLPS, thereby providing a foundation for future drug development.

2. Methods

2.1. Computational workflow

An integrated computational pipeline was constructed combining ensemble machine learning-based QSAR, ensemble molecular docking against MDM2 and MDM4, molecular dynamics simulations, network pharmacology and ADMET profiling. The pipeline was applied to a curated drug-like subset of the ZINC20 database to prioritize candidate dual MDM2/MDM4 inhibitors for dedifferentiated liposarcoma (DDLPS). Network pharmacology, informed by DDLPS patient transcriptomic data, predicted polypharmacological potential against disease-overexpressed genes.

2.2. Data curation and preparation

Compound bioactivity data against MDM2 (ChEMBL target ID ChEMBL5023) were retrieved from ChEMBL version 32 [42]. Only entries with reported IC_{50} values were retained. For duplicate canonical SMILES, the entry with the lowest (most potent) IC_{50} was kept. Invalid structures were excluded using RDKit-based validation (Landrum, 2016) [44]. IC_{50} values (nM) were converted to $pIC_{50} = -\log_{10}(IC_{50} \times 10^{-9})$. This transformation is widely adopted to normalize activity distributions and improve regression stability (Tropsha, 2010; Cherkasov et al., 2014) [45,41]. Outliers were removed using the interquartile range (IQR) method with a multiplier of 2.5: lower bound = $Q1 - 2.5 \times IQR$, upper bound = $Q3 + 2.5 \times IQR$, where $Q1$ and $Q3$ are the 0.05 and 0.95 quantiles of pIC_{50} , respectively. The final curated dataset contained 4,154 unique compounds.

2.3. Molecular representation and feature engineering

Each compound was represented by extended-connectivity fingerprints (ECFP4) with radius 2 and 1024 bits, generated using RDKit version 2020.09. ECFP4 is a circular topological fingerprint that encodes the presence of local structural features within a defined atomic neighborhood, making it widely used for capturing structure-activity relationships in QSAR modelling [46]. Features with variance below 0.01 were removed using VarianceThreshold from scikit [47]-learn. From the remaining features, mutual information regression between each feature and the pIC_{50} values was computed, and the top 200 features with the highest mutual information scores were selected for model building.

2.4. Model development and optimization

The dataset was split into training (70%), test (15%) and validation (15%) sets using a stratified split based on pIC_{50} quantiles (via `train_test_split` with `test_size=0.3`, then splitting the resulting 30% into equal halves). All models were trained only on the training set. Features were standardised to zero mean and unit variance using StandardScaler. Five individual regression algorithms (XGBoost [48], LightGBM [49], Random Forest [50], Gradient Boosting [52] and Extra Trees [51]) were implemented, with hyperparameters optimised via randomised search and five-fold cross-validation on the training set. These tree-based ensemble methods are well suited to capturing non-linear structure-activity relationships. A voting ensemble regressor was also constructed using the first four models (XGBoost, Random Forest, LightGBM, Gradient Boosting) with equal weighting. To ensure reproducibility, a fixed random seed of 42 was used for all stochastic operations, including data splitting, model initialization, and cross-validation folds.

2.5. Validation metrics and model selection

All six models were evaluated on the test set. The following metrics were calculated: coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), concordance correlation coefficient (CCC), Pearson and Spearman correlation coefficients, mean absolute percentage error (MAPE), and three variants of predictive squared correlation coefficients: Q^2_{F1} , Q^2_{F2} and Q^2_{F3} . Five-fold cross-validation predictions on the training set were used to compute Q^2_{CV} . Additionally, the slopes k and k' from regression of observed versus predicted values were calculated, together with the squared correlation coefficients through the origin (R_0^2 and $R_0'^2$).

Y-randomization was performed to assess whether the observed predictive performance could arise from chance correlations. The target variable was randomly permuted, and the models were rebuilt repeatedly. A substantial decrease in predictive performance relative to the original model was taken as evidence of meaningful structure-activity relationships.

To demonstrate that the AI approach offers added value over simpler baselines, we compared the performance of the Random Forest model against a simple linear regression model and a k-nearest neighbors (k-NN) regressor ($k=5$) using the same feature set. The linear regression achieved a test R^2 of 0.52 ± 0.03 , while k-NN achieved 0.51 ± 0.04 , both substantially lower than the Random Forest model ($R^2 = 0.76$). This confirms that the ensemble tree-based approach captures non-linear structure-activity relationships that simpler models fail to identify.

2.6. Applicability domain and virtual screening

The applicability domain of the best model was defined using the leverage approach, commonly visualized in a Williams plot. The warning threshold was set to $h^* = 3(p + 1)/n$, where p is the number of descriptors and n is the number of training samples. A compound was considered within the applicability domain if $h \leq h^*$; otherwise its prediction was flagged as unreliable. Standardized residuals were also evaluated, with compounds falling outside ± 3 standard deviations considered outliers.

A curated drug-like subset of the ZINC20 database was assembled for virtual screening. The subset was filtered to retain compounds compliant with Lipinski's rule of five (molecular weight ≤ 500 Da, calculated $\text{LogP} \leq 5$), commercially available from in-stock catalogs, and prepared in ready-to-dock three-dimensional formats. The resulting library comprised approximately 9.4 million unique substances. The optimized QSAR models were applied to this subset to predict biological activity. Predictions included pIC_{50} values, back-transformed IC_{50} estimates and applicability domain classification. Compounds were ranked by predicted activity to facilitate the identification of potential MDM2/MDM4 inhibitors. To manage memory, the compound file was read in chunks of 50,000 entries. For each compound, ECFP4 fingerprints were generated, followed by the identical feature selection and scaling transformations applied to the training data. To prevent data leakage, the feature selection and scaling transformations were fitted exclusively on the training data and then applied to the validation, test, and screening sets. No information from the screening library was used in model training or feature selection. Compounds with a predicted pIC_{50} of 8.0

or higher (equivalent to an $IC_{50} \leq 10$ nM) that fell within the applicability domain were selected as candidate MDM2 inhibitors for subsequent ensemble molecular docking against MDM2 and MDM4.

2.7. Ensemble molecular docking

2.7.1. Docking protocol validation

The docking protocol was validated using two complementary approaches. First, for each co-crystallised ligand present in the MDM2 and MDM4 structures that were included in the ensemble (regardless of whether binding affinity data were available), the ligand was extracted, prepared and re-docked into its original binding pocket using PyRx v0.8 [54] with an exhaustiveness of 128 [53]. For each receptor, a grid box of $25 \text{ \AA} \times 25 \text{ \AA} \times 25 \text{ \AA}$ was centred on the co-crystallised ligand. The root-mean-square deviation (RMSD) between the re-docked pose and the crystallographic ligand position was calculated; successful validation required RMSD below $2.0 \text{ \AA}$ for all structures.

Second, for the subset of co-crystallised ligands with available experimental binding data (K_i , K_d or IC_{50}) available in the Protein Data Bank, the predicted binding free energy (ΔG) from docking was compared with the experimental ΔG derived from the equilibrium constant using $\Delta G = RT \ln(K)$ ($T = 298 \text{ K}$, $R = 1.987 \text{ cal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$). Overlap between the predicted and experimental ΔG ranges was considered evidence of predictive accuracy.

2.7.2. Protein preparation

For MDM2, an ensemble of 23 crystal structures with resolution $\leq 3.0 \text{ \AA}$ was assembled from the Protein Data Bank (PDB IDs: 1RV1, 1T4E, 1T4F, 1YCR, 2AXI, 2GV2, 3EQS, 3G03, 3IUX, 3JZK, 3JZR, 3JZS, 3LBK, 3LBL, 3LNJ, 3LNZ, 3TJ2, 3TPX, 3TU1, 3V3B, 3VZV, 3W69, 4DIJ). For MDM4, an ensemble of 24 crystal structures (PDB IDs: 3DAB, 3EQY, 3FDO, 3FE7, 3FEA, 3JZO, 3JZP, 3JZQ, 3LBJ, 3U15, 4RXZ, 5UML, 5VK1, 6Q9Q, 6Q9S, 6Q9U, 6Q9W, 6Q9Y, 7C3Q, 7C3Y, 7C44, 7KJN, 8HDG and 8IA5) was used. All protein structures were prepared using UCSF Chimera v1.17 [55]: water and heteroatoms were removed, hydrogens added at pH 7.4, Gasteiger charges assigned, and energy minimized with the AMBER force field. Prepared structures were converted to Pdbqt format using PyRx v0.8.

2.7.3. Ligand preparation

The top-ranking compounds (predicted $pIC_{50} > 8.0$) from the virtual screening were prepared using LigPrep (Maestro) [56]. Tautomers, stereoisomers and protonation states at $\text{pH } 7.4 \pm 0.5$ were generated, and energy minimization was performed with the OPLS4 force field. All ligands were converted to pdbqt format.

2.7.4. Docking protocol and analysis

Compounds with a predicted pIC_{50} of 8.0 or higher (equivalent to an $IC_{50} \leq 10$ nM) that fell within the applicability domain were selected as candidate MDM2 inhibitors for subsequent ensemble molecular docking against MDM2 and MDM4. As controls, 18 MDM2/MDM4 bound ligands (MDM2: 0R3, 1GI, 1MT, 4SS, 6SK, HRH, IMZ, K23, NUT, TUZ, YIN, MDM4: 03M, HUE, HRE, HRT, HRQ, HRN, NUT201, and WW8) and 30 low-potency compounds (predicted $pIC_{50} < 5$ by the best model) were also selected and subjected to ligand docking. Each ligand was docked independently into each receptor of the MDM2 ensemble and the MDM4 ensemble, generating ten poses per run. Three independent docking runs with

different random seeds were performed to ensure sampling diversity. For each ligand, the binding affinity (ΔG) across the MDM2 ensemble and across the MDM4 ensemble was recorded. Protein-ligand interaction analysis was performed utilizing Discovery Studio Visualizer v2024.

2.8. Molecular dynamics simulations

For the top-ranked compounds, molecular dynamics (MD) simulations were performed for two systems: MDM2 (PDB: 3JZK) and MDM4 (PDB: 8IA5). Each complex was prepared using the CHARMM-GUI Solution Builder [57], solvated in a cubic TIP3P water box with a 10 Å buffer, neutralised with 0.15 M NaCl, and assigned the CHARMM36 force field [58]. Systems were energy-minimised using the steepest descent algorithm (10,000 steps) with positional restraints ($1,000 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$) on protein and ligand heavy atoms. Equilibration comprised a 75 ps NVT simulation with the V-rescale thermostat (300 K) followed by a 5 ns NPT simulation with the Parrinello-Rahman barostat (1 atm), both with positional restraints. Production simulations were run for 500 ns with a 2 fs time step, using the LINCS algorithm to constrain hydrogen bonds and the Particle Mesh Ewald method for long-range electrostatics. All simulations were performed with GROMACS 2023.1 [59]. Trajectory analyses were performed with GROMACS utilities and in-house Python scripts. Standard metrics included root-mean-square deviation (RMSD) of protein C α atoms and ligand heavy atoms, root-mean-square fluctuation (RMSF) of C α atoms, radius of gyration (Rg) and solvent-accessible surface area (SASA). Principal component analysis of C α atomic fluctuations was performed, and the Gibbs free energy landscape was estimated.

2.9. MM/GBSA binding free energy calculations

Binding free energies were estimated using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method as implemented in gmx_MMPBSA v1.6.3 [43]. For each of the simulated compounds, the last well-equilibrated section of the 500 ns trajectory (3500-4500 ns, 10,000 frames at 10 ps intervals) were used. Per-residue decomposition was also performed for residues within 6 Å of the ligand.

2.10. Network pharmacology with DDLPS Transcriptomics

To define a DDLPS-specific gene expression signature, the whole-transcript microarray dataset GSE21122 (Affymetrix Human Genome U133A array) was retrieved from the Gene Expression Omnibus. The dataset comprises 158 soft-tissue sarcoma specimens (46 dedifferentiated liposarcoma, 26 leiomyosarcoma, 34 myxofibrosarcoma/pleomorphic malignant fibrous histiocytoma, 20 myxoid/round cell liposarcoma, 23 pleomorphic liposarcoma) and 9 normal adipose tissue samples (Barretina et al. [8]). All 46 DDLPS tumours (GSM528276-GSM528321) and all nine normal fat controls (GSM528425-GSM528433) were selected based on histological annotations. Differential expression analysis was performed using GEO2R, which implements the limma package with voom precision weights. Raw CEL files were normalised using the robust multi-array average (RMA) algorithm, and probe-level data were collapsed to gene symbols using the manufacturer's annotation (GPL96). A linear model was fitted comparing DDLPS versus normal fat, and moderated t-statistics were computed. P-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) procedure. Genes with an FDR-adjusted $P < 0.05$ and an absolute $\log_2$ fold change ≥ 1 (i.e., ≤ -1 or $\geq +1$) were considered significantly differentially expressed.

Separately, a curated list of genes associated with dedifferentiated liposarcoma was retrieved from the GeneCards database (<https://www.genecards.org>) using the search term "dedifferentiated liposarcoma" [62]. The intersection between the differentially expressed genes from GSE21122 and the GeneCards-derived DDLPS gene set was taken as the final DDLPS-relevant gene set. Potential protein targets of the selected compounds were predicted using SwissTargetPrediction (<http://www.swisstargetprediction.ch>) [61] with the Homo sapiens proteome. The predicted target lists for the two compounds were merged and duplicate entries removed. The intersection between this merged compound-predicted target set and the final DDLPS-relevant gene set was then identified. The intersecting genes were used as input for protein-protein interaction (PPI) network construction using the STRING database v12.0 (<https://string-db.org>) [60]. A confidence score threshold of 0.7 was applied, and all other parameters were kept at default values. Functional enrichment analysis for KEGG pathways, Gene Ontology (GO) biological processes, and Disease Ontology terms was performed within STRING [60] using the same confidence threshold, with an FDR < 0.05 considered statistically significant.

2.11. ADMET profiling

Absorption, distribution, metabolism, excretion and toxicity (ADMET) properties were predicted for the selected compounds using the machine-learning-based tools SwissADME [63] and ADMETlab 3.0 [64]. Key parameters assessed included molecular weight, topological polar surface area, aqueous solubility (LogS), human intestinal absorption, plasma protein binding, blood-brain barrier penetration, CYP450 inhibition (CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP3A4), P-glycoprotein substrate status and hERG blockade risk. Thresholds for drug-likeness were interpreted according to established guidelines.

3.0. Results

3.1. Machine Learning-QSAR

3.1.1 Dataset preparation and model development

From ChEMBL version 32 [42], 5,418 compounds with reported IC_{50} values against MDM2 were retrieved. After removing duplicate SMILES and filtering outliers using a $2.5 \times IQR$ criterion, 4,154 unique compounds remained. The pIC_{50} values ranged from 3.60 to 10.40 (mean 7.75 ± 1.33). Compounds were encoded as ECFP4 fingerprints (radius 2, 1024 bits). Variance filtering (threshold 0.01) removed near-constant features, and the top 200 features were selected by mutual information regression. The data were partitioned into training (70%, 2,907 compounds), test (15%, 623 compounds), and validation (15%, 624 compounds) sets using a stratified split based on pIC_{50} quantiles.

Six regression models (XGBoost, LightGBM, Random Forest, Gradient Boosting, Extra Trees, and a voting ensemble) were trained and evaluated on the MDM2 bioactivity dataset. **Table 1** presents their performance across multiple validation metrics, grouped into four categories: goodness of fit (training R^2 , test R^2 , ΔR^2), prediction error (RMSE, MAE, MAPE), internal and external robustness (Q^2CV , Q^2F1 , Q^2F2 , Q^2F3), and agreement and bias (CCC, Pearson r , Spearman ρ , slopes k and k' , squared correlations through the origin R_0^2 and $R_0'^2$).

All models yielded test R^2 values between 0.7385 and 0.7628. Random Forest achieved the highest test R^2 (0.7628), the lowest test RMSE (0.6399), and the lowest test MAE (0.4479). Its concordance correlation coefficient (CCC) of 0.8670 was the highest among the models, indicating greater precision and accuracy. The difference between training and test R^2 ($\Delta R^2 = 0.1158$) was small, suggesting minimal overfitting. The cross-validated Q^2 (0.7212) was comparable to the test R^2 , and the external predictivity metrics Q^2F1 , Q^2F2 , and Q^2F3 were 0.8682, 0.7629, and 0.7628, respectively, each exceeding the recommended threshold of 0.6. The slope k of observed versus predicted values was 0.9999, and the intercept-free squared correlations R_0^2 and $R_0'^2$ were both 0.9933, meeting the Tropsha criteria for externally validated QSAR models.

The voting ensemble achieved a test R^2 of 0.7576, slightly below that of Random Forest. Although its training R^2 was higher (0.8956), its ΔR^2 was larger (0.1380), and its slope k (0.9780) deviated more from the ideal value of 1, indicating greater overfitting. LightGBM (test $R^2 = 0.7509$) and ExtraTrees (test $R^2 = 0.7465$) exhibited larger ΔR^2 (0.1275 and 0.1649) and higher test RMSE (0.6558 and 0.6615) than Random Forest. ExtraTrees showed the highest training R^2 (0.9114) and the lowest Q^2CV (0.7021). GradientBoosting (test $R^2 = 0.7445$) and XGBoost (test $R^2 = 0.7385$) yielded the lowest test R^2 and the largest ΔR^2 (0.1350 and 0.1809), with XGBoost also recording the lowest Q^2CV (0.6855). All models yielded CCC values above 0.85, Pearson correlation coefficients exceeding 0.86, and Spearman rank correlation coefficients above 0.85, demonstrating strong linear and monotonic agreement between predicted and

observed pIC₅₀ values. Mean absolute percentage error (MAPE) values ranged from 6.35% to 6.58%, confirming that prediction errors were consistently low across the six models.

Table 1: Validation metrics for MDM2 pIC₅₀ prediction models

Model	Random Forest	LightGBM	Extra Trees	Gradient Boosting	XGBoost	Voting Ensemble
R^2_{train}	0.88	0.88	0.91	0.88	0.92	0.90
R^2_{test}	0.76	0.75	0.75	0.74	0.74	0.76
$\Delta R^2_{(T-T)}$	0.12	0.13	0.16	0.13	0.18	0.14
R^2_{val}	0.67	0.67	0.64	0.66	0.63	0.66
Q^2_{CV}	0.72	0.72	0.70	0.72	0.69	0.72
RMSE _{train}	0.46	0.46	0.40	0.46	0.38	0.43
RMSE _{test}	0.64	0.66	0.66	0.66	0.67	0.65
RMSE _{val}	0.78	0.78	0.80	0.79	0.82	0.78
MAE _{train}	0.31	0.31	0.24	0.32	0.23	0.28
MAE _{test}	0.45	0.46	0.45	0.46	0.45	0.45
MAE _{val}	0.53	0.52	0.54	0.52	0.54	0.52
Q^2_{F1}	0.87	0.85	0.82	0.85	0.80	0.85
Q^2_{F2}	0.76	0.75	0.75	0.74	0.74	0.76
Q^2_{F3}	0.76	0.75	0.75	0.74	0.74	0.76
CCC	0.87	0.86	0.86	0.86	0.86	0.87
K	1.00	0.96	0.95	0.97	0.94	0.98
k'	-0.04	0.21	0.34	0.16	0.44	0.12
R_o^2	0.99	0.99	0.99	0.99	0.99	0.99
R'_o^2	0.99	0.99	0.99	0.99	0.99	0.99
Pearson (r)	0.87	0.87	0.87	0.86	0.86	0.87
Spearman (rho)	0.86	0.86	0.85	0.86	0.85	0.86
MAPE	6.36	6.49	6.40	6.58	6.43	6.35

R^2 , coefficient of determination; $\Delta R^2_{(T-T)}$, difference between training and test R^2 ; Q^2_{CV} , cross-validated R^2 ; RMSE, root-mean-square error; MAE, mean absolute error; Q^2_{F1} , Q^2_{F2} , Q^2_{F3} , external predictivity metrics; CCC, concordance correlation coefficient; k and k' , slopes of regression lines; R_o^2 and R'_o^2 , squared correlation coefficients through the origin; MAPE, mean absolute percentage error.

3.1.2 Model validation and Y-randomization

To rule out chance correlations, y-randomization testing was conducted. The response vector (pIC_{50}) was randomly permuted 50 times while the descriptor matrix was held fixed; each model was retrained using the same hyperparameters. For every model, all randomized R^2 values were negative, indicating performance worse than a trivial mean predictor ($R^2 = 0$). The random forest model recorded a mean randomized R^2 of -0.110 ± 0.009 , compared with an original R^2 of 0.714, and every permutation yielded a value far below the commonly accepted threshold for chance correlation (0.05). The consistently negative R^2 across all permutations, together with the tight standard deviation, demonstrates that randomized responses produce stable, uniformly poor performance. This provides strong evidence that the models capture genuine structure-activity relationships and that the descriptor set does not introduce structural bias.

The random forest model was externally validated on 45 MDM2 inhibitors with literature-reported IC_{50} values. It achieved an R^2 of 0.79, a root-mean-square error of 0.78 pIC_{50} units, a mean absolute error of 0.52 pIC_{50} units, and a geometric mean fold error in IC_{50} of 2.01, meaning predictions are typically within twofold of experiment. These results indicate strong generalisation to unseen compounds and support reliable prioritisation of active molecules for subsequent screening.

3.1.3 Applicability domain and virtual screening

The applicability domain (AD) of the Random Forest model was defined using the leverage approach. The warning leverage threshold was set to $h^* = 3(p + 1)/n$, where $p = 200$ (selected features) and $n = 2,907$ (training compounds), yielding $h^* = 0.207$. For the 623 test-set compounds, residuals ranged from -4.55 to +2.17 log units, with a mean of -0.04 and a standard deviation of 0.64. Standardised residuals fell predominantly within ± 2.5 , indicating the absence of systematic bias. A small fraction of test compounds (1.6%; 10 out of 623) exhibited standardised residuals exceeding ± 3 . All ten of these high-residual compounds had leverage values below h^* , confirming that they lay within the descriptor-based AD yet were poorly predicted, underscoring that the leverage-based AD defines the chemical space in which predictions are structurally reliable but does not guarantee accuracy for compounds with response values in under-represented regions of the activity spectrum. Only 9 compounds (1.4%) exceeded the leverage threshold, placing them outside the model's AD (**Figs. S1, S2**).

The Random Forest model was retrained on the complete curated dataset of 4,154 compounds and applied to screen a curated drug-like subset of the ZINC20 database (approximately 9.4 million compounds) filtered for compliance with Lipinski's rule of five (molecular weight ≤ 500 Da, $\text{LogP} \leq 5$) and prepared in ready-to-dock three-dimensional formats. Screening yielded 402 high-potency candidates, with predicted pIC_{50} values ranging from 8.00 to 8.63 (mean 8.17 ± 0.17) and corresponding IC_{50} values from 2.36 nM to 9.96 nM (mean 7.25 nM, median 8.09 nM). Among these, only 89 compounds (22.1%) had leverage values $\leq h^*$ (0.207) and were therefore considered reliable (**Table S1**). The remaining 313 compounds (77.9%) fell outside the applicability domain and were excluded from further consideration. The refined set of

AD-compliant high-potency hits was advanced to ensemble docking against MDM2 and MDM4 to predict their binding affinities.

3.2. 0. Molecular docking

3.2.1. Validation of docking protocol

The docking protocol was validated by redocking co-crystallized ligands into their cognate binding pockets across the 23 MDM2 and 24 MDM4 structures used for ensemble docking. For every structure with a bound ligand, the root-mean-square deviation (RMSD) between the redocked pose and the crystallographic pose remained below 2.0 Å (**Table S2**), indicating that the protocol reproduces experimental binding modes with accuracy suitable for ensemble docking. This result is consistent with previous AutoDock Vina benchmarks using an exhaustiveness of 128 [53].

Agreement between docking scores and experimentally derived binding free energies ($\Delta G = RT \ln K$, $T = 298\text{ K}$) was assessed across multiple chemotypes (**Table S2**). For MDM2 ligands, scoring ranges encompassed or approached experimental values: spiro-oxindole MI6 (RMSD 0.08–0.99 Å, scores -11.98 to -7.2 kcal mol⁻¹, experimental ΔG -11.62 kcal mol⁻¹), isoindolinone K23 (RMSD 0.20–1.18 Å, -9.9 to -8.1 kcal mol⁻¹, experimental ΔG -8.23 kcal mol⁻¹), imidazoline IMZ (RMSD 0.16–0.21 Å, -9.6 to -7.4 kcal mol⁻¹, ΔG -9.34 kcal mol⁻¹), TJ2 (RMSD 0.30–1.19 Å, -8.8 to -7.8 kcal mol⁻¹, ΔG -8.72 kcal mol⁻¹), and nutlin-3a across four MDM2 structures (RMSD 0.08–1.12 Å, -7.4 to -5.8 kcal mol⁻¹, experimental ΔG -7.90 to -5.67 kcal mol⁻¹). For MDM4, indolyl hydantoin 03M gave scores from -10.2 to -7.0 kcal mol⁻¹ (experimental ΔG -10.07 kcal mol⁻¹), while HRQ (PDB 6Q9Y) yielded -7.7 to -6.1 kcal mol⁻¹ (experimental ΔG -7.41 kcal mol⁻¹). Although score ranges covered experimental values, distributions skewed toward less favorable energies, notably for 03M and MI6, likely reflecting limitations of the AutoDock Vina scoring function in capturing subtle energetic contributions. Overall, the docking protocol recovers crystallographic poses and provides binding-energy estimates in useful agreement with experiment, supporting its application to the ensemble-based virtual screening.

3.2.2. Ensemble docking

Ensemble docking was performed for 89 AD-compliant, high-potency hits ($\text{pIC}_{50} \geq 8$), 11 MDM2-bound ligands, 8 MDM4-bound ligands, and 30 low-potency compounds (predicted $\text{pIC}_{50} < 5$) against 23 MDM2 and 24 MDM4 crystal structures. Full docking results are provided in the Supplementary Information (**Tables S3, S4**).

Against MDM2, high-potency hits recorded a mean D-score of -7.21 ± 0.58 kcal mol⁻¹, more favourable than low-potency compounds (-3.73 ± 0.44 kcal mol⁻¹; t-test $p < 10^{-43}$), indicating discrimination between active and inactive compounds. Known MDM2 inhibitors achieved a mean D-score of -7.88 ± 0.49 kcal

mol⁻¹, lower than both high-potency hits ($p = 8.9 \times 10^{-4}$) and low-potency compounds ($p = 1.8 \times 10^{-14}$). The best-performing known inhibitor, K23, exhibited a mean D-score of -8.60 kcal mol⁻¹. Two high-potency hits (ZINC9412756 and ZINC20724474) showed mean D-scores of -9.27 ± 0.38 and -9.10 ± 0.38 kcal mol⁻¹, respectively, more negative than K23. One-way ANOVA across the three groups was significant ($F = 499.2$, $p < 10^{-60}$).

Against MDM4, the same hits gave a mean D-score of -6.52 ± 0.57 kcal mol⁻¹, more favourable than low-potency compounds (-3.59 ± 0.42 kcal mol⁻¹; $p < 10^{-40}$). The eight known MDM4 inhibitors had a mean D-score of -7.11 ± 0.45 kcal mol⁻¹, lower than low-potency compounds ($p < 10^{-16}$). The best-performing known MDM4 inhibitor, HRN, achieved a mean D-score of -7.80 kcal mol⁻¹. ZINC9412756 and ZINC20724474 produced mean D-scores of -9.27 and -9.20 kcal mol⁻¹ on MDM4, both more favourable than HRN. One-way ANOVA was significant ($F = 364.0$, $p < 10^{-52}$). Mean docking scores for MDM2 and MDM4 showed a positive correlation (Pearson $r = 0.924$), consistent with the structural similarity of their p53-binding clefts. The two compounds with mean D-scores more favourable than all bound ligands were selected for further evaluation as predicted dual inhibitors.

3.2.3. Protein-ligand interaction analysis

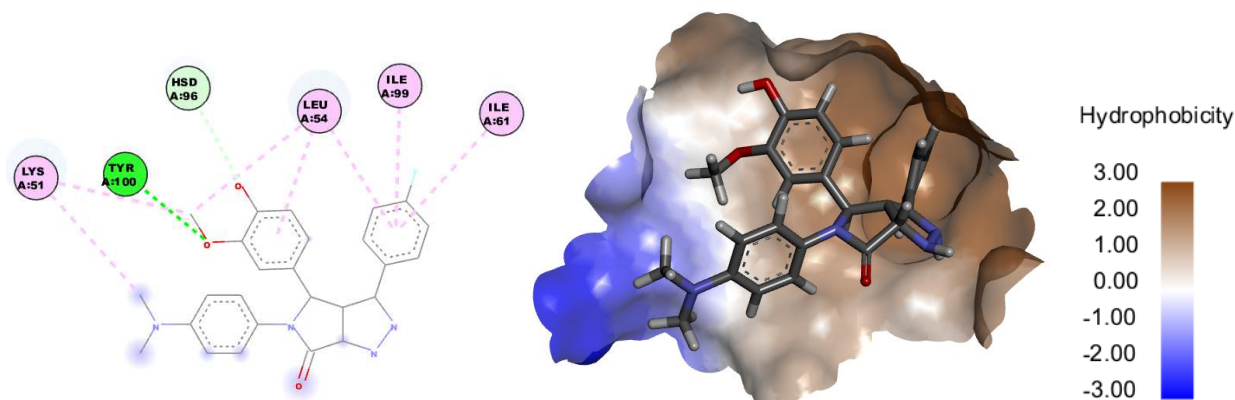
To understand the molecular basis for the predicted binding affinities of ZINC20724474 and ZINC9412756, we examined their docked poses within the p53-binding clefts of MDM2 and MDM4 alongside known inhibitors (**Fig. 2, Table S5; Fig. 3, Table S6**). Interaction distances (Å) and donor-hydrogen-acceptor angles (°) for hydrogen bonds are reported to assess geometric favorability.

In the MDM2 binding pocket, co-crystallized inhibitors (0R3, 1GI, 1MT, 4SS, 6SK, HRH, IMZ, K23, NUT, TUZ, YIN) occupied hydrophobic subpockets defined by Leu54, Ile61, Met62, Val93, Ile99, and His96, forming extensive alkyl or π -alkyl contacts (3.5–5.5 Å). Several inhibitors also engaged the protein through conventional hydrogen bonds. 1MT formed two such bonds with the backbone carbonyl of Leu54 (2.11 Å, 152°) and N ϵ 2 of His96 (2.79 Å, 124°). 6SK established three hydrogen bonds: with Leu54 O (2.49 Å, 141°), Val93 O (2.03 Å, 164°), and His96 N ϵ 2 (2.70 Å, 104°). The highest-affinity control, K23 ($K_i = 916$ nM, experimental $\Delta G \approx -8.23$ kcal/mol), contributed a single hydrogen bond to Leu54 O (1.98 Å, 125°) and a face-to-face π - π stack with His96 (centroid distance 3.54 Å, interplanar angle $\sim 7.6^\circ$); its chlorine atoms formed alkyl contacts with Leu57, Ile61, Val75, and Val93 (4.09–4.99 Å). Other controls relied predominantly on hydrophobic packing and carbon-hydrogen bonds.

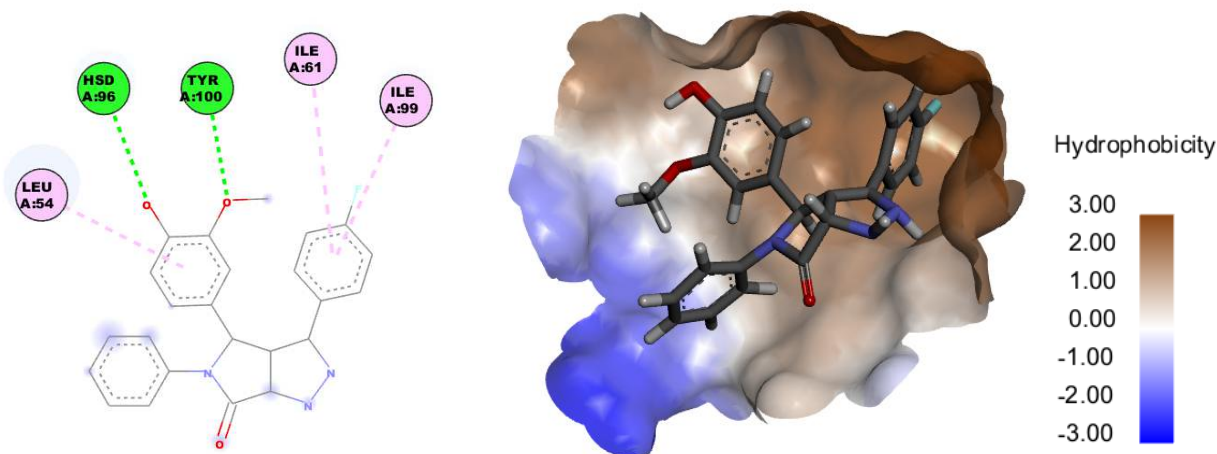
ZINC20724474 formed a short, near-linear conventional hydrogen bond between its carbonyl oxygen O1 and the Tyr100 hydroxyl (1.82 Å, 163°), together with a carbon-hydrogen bond between His96 HD2 and ligand O2 (2.74 Å, 125°). The hydrophobic interface comprised alkyl contacts of C11 with Lys51 and Leu54, C19 with Lys51, and π -alkyl contacts with Leu54 (4.26, 4.94 Å), Ile61 (5.22 Å), and Ile99 (4.82 Å). This combination of a Tyr100-anchored hydrogen bond and a dense hydrophobic network is consistent with its

predicted binding affinity of $-9.27 \text{ kcal mol}^{-1}$. The binding mode shares features with 1MT and 6SK, both of which engage His96 and Leu54, but neither forms a hydrogen bond with Tyr100. ZINC20724474 thus extends the 1MT/6SK binding motif by incorporating a Tyr100 anchor, resulting in a more extensive polar interaction network and improved predicted affinity (-9.27 vs. -7.8 to -8.6 kcal/mol for controls).

ZINC9412756 formed two conventional hydrogen bonds: His96 HD1 to O2 ($2.49 \text{ \AA}$, 117°) and Tyr100 HH to O1 ($2.09 \text{ \AA}$, 139°), with additional π -alkyl contacts to Leu54 ($4.32 \text{ \AA}$), Ile61 ($5.21 \text{ \AA}$), and Ile99 ($5.42 \text{ \AA}$). Although its Tyr100 interaction is longer and less linear than that of ZINC20724474, the dual hydrogen-bond motif with both His96 and Tyr100 is unique among controls. The dual His96-Tyr100 engagement creates a polar clamp that stabilizes the ligand in a manner not observed in any co-crystallized inhibitor. This polar architecture, combined with efficient hydrophobic packing, yields a predicted binding affinity of $-9.20 \text{ kcal mol}^{-1}$. The superior docking scores of candidate compounds relative to controls can be attributed to two complementary features: the Tyr100 hydrogen bond, which recapitulates a key p53 interaction not fully exploited by existing inhibitors, and optimized hydrophobic packing within the Leu54-Ile61-Ile99 subpocket. These features suggest that the ZINC compounds occupy a subtly different region of the MDM2 binding site compared to nutlin-class inhibitors, potentially offering a distinct chemotype for MDM2 inhibition.



(A) ZINC20724474



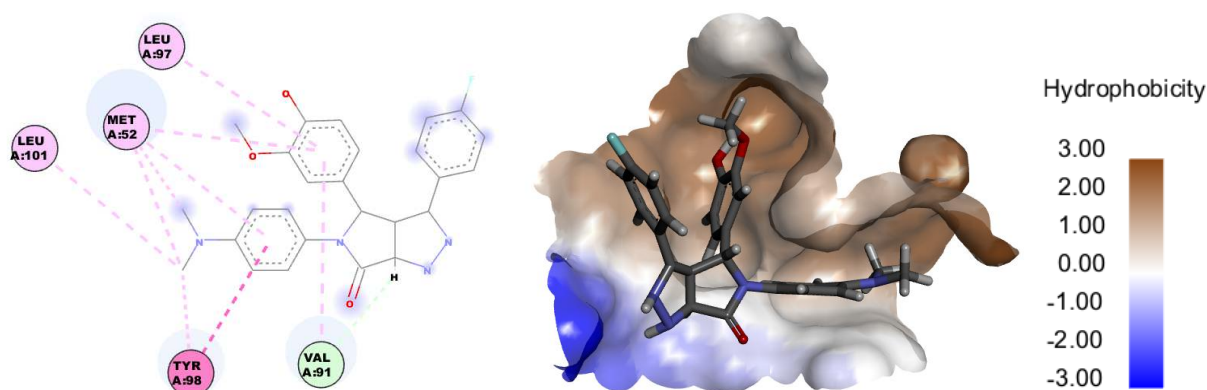
(B) ZINC9412756

Fig. 2: MDM2-Ligand Interaction Maps. Right panels show 3D surface representations of the binding pocket illustrating spatial accommodation of ligands and interactions with surrounding residues. Left panels show 2D interaction diagrams depicting hydrogen bonds, hydrophobic interactions, and other key contacts between ligands and active site residues for (A) **ZINC20724474** and (B) **ZINC9412756**

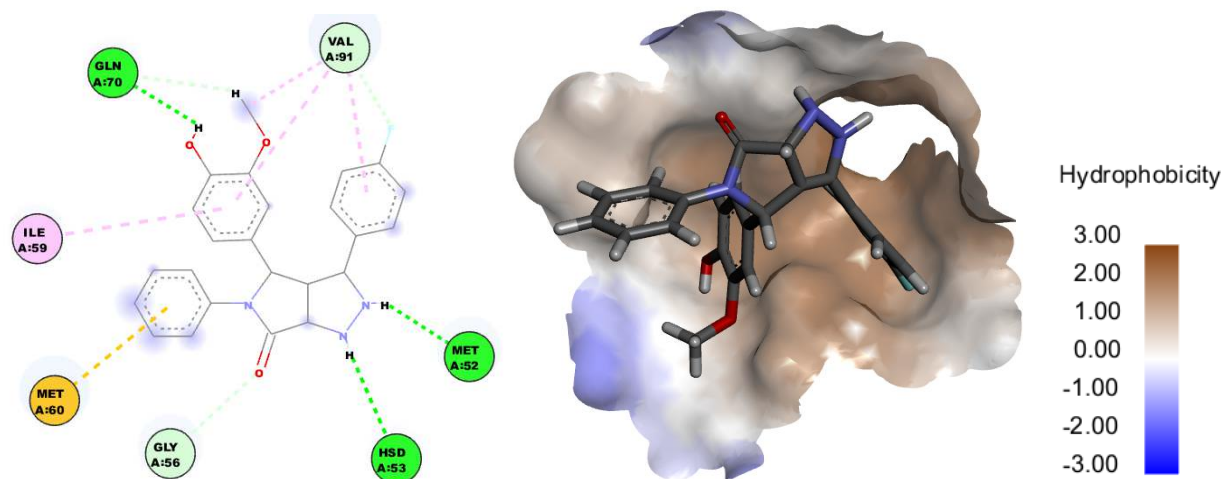
In the MDM4 binding pocket, co-crystallized inhibitors (03M, HUE, HRE, HRT, HRQ, HRN, NUT201, WW8) occupied hydrophobic subpockets defined by Met52, Leu55, Ile59, Leu97, Val91, Tyr98, and Pro94. All ligands formed at least one π - π or π - σ interaction with Tyr98 or Val91, and six established at least one conventional hydrogen bond. The strongest such bonds were observed for HUE (two bonds to the backbone carbonyl of Met52: 1.86 Å, 160°; and 2.12 Å, 177°) and WW8 (2.07 Å, 144°). Weaker hydrogen bonds were seen for HRT with Gln70 (2.20 Å, 126°) and HRQ with Thr107 (2.24 Å, 122°). π -Sulfur interactions with Met52, present in 03M, HRQ, and HRN (4.09–4.38 Å), provided additional stabilization. NUT201 lacked any conventional hydrogen bond, relying instead on a single carbon-hydrogen bond (2.70 Å) and a π - σ contact, consistent with its moderate predicted binding affinity (-7.29 kcal mol $^{-1}$).

ZINC20724474 established a predominantly hydrophobic recognition mode in MDM4, with no conventional hydrogen bond formed. Instead, the ligand is anchored by a carbon-hydrogen bond between the C25-H group and the backbone carbonyl of Val91 (2.5 Å) and a face-to-face π - π stacking interaction with Tyr98 (4.74 Å). Extensive hydrophobic packing further stabilizes the complex through alkyl and π -alkyl contacts with Met52, Leu101, Val91, and Leu97. Despite the absence of a canonical hydrogen bond, the predicted binding free energy (-9.27 kcal mol $^{-1}$) is more favourable than that of any co-crystallized MDM4 control. The binding mode closely resembles that of 03M (IC $_{50}$ = 41 nM, experimental $\Delta G \approx -10.07$ kcal/mol), sharing the Tyr98-Met52-Leu101 triad but replacing the Thr107 H-bond with a C-H $\cdots$ Val91 interaction. This substitution retains the polar anchor while engaging a different pocket region (Val91 vs. Thr107). The docking score (-9.27 kcal/mol) falls within the range observed for 03M across the ensemble (-10.2 to -7.0 kcal/mol), suggesting that this hydrophobic-dominated strategy can achieve high predicted affinity.

ZINC9412756 established the most extensive polar interaction network of any MDM4 ligand evaluated, forming three conventional hydrogen bonds (more than all co-crystallized controls). The ligand donates N-H $\cdots$ O bonds to the backbone carbonyls of Met52 (2.08 Å, 143°), His53 (2.88 Å, 138°), and Gln70 (2.21 Å, 134°). The polar network is further strengthened by two carbon-hydrogen bonds: from the amide nitrogen-bound hydrogen (H9) to the Gln70 carbonyl (2.90 Å) and from Gly56 HA to the ligand methoxy oxygen (2.62 Å, 130°). A C-H $\cdots$ F interaction with Val91 HA (2.99 Å, 117°, 138°) contributes additional polar stabilization. The ligand is further embedded through a π -sulfur contact with Met60 (5.43 Å) and π -alkyl contacts with Ile59 and Val91. The predicted binding affinity (-9.20 kcal mol $^{-1}$) is consistent with this extended polar network. The binding mode shares features with HUE (two hydrogen bonds to Met52) but expands the polar network by additionally engaging His53 and Gln70, neither targeted by any reference ligand. This multi-point polar engagement, combined with hydrophobic packing, positions ZINC9412756 as a structurally distinct chemotype from nutlin-like inhibitors, which rely primarily on chlorophenyl-leucine pocket penetration. The binding mode demonstrates that a hydrogen-bond-rich motif can achieve high predicted affinity for MDM4, complementing the hydrophobic strategy exemplified by 03M and expanding the repertoire of viable chemotypes for this target.



ZINC20724474



ZINC9412756

Fig. 3: MDM4-Ligand Interaction Maps. Right panels show 3D surface representations of the binding pocket illustrating spatial accommodation of ligands and interactions with surrounding residues. Left panels show 2D interaction diagrams depicting hydrogen bonds, hydrophobic interactions, and other key contacts between ligands and active site residues for (A) **ZINC20724474** and (B) **ZINC9412756**.

3.2.4: Structural Novelty Assessment

To assess the structural novelty of the identified candidate compounds, we performed a systematic similarity analysis against all known MDM2 and MDM4 inhibitors from ChEMBL (version 32). The MDM2 reference set contained 4,154 unique compounds and the MDM4 set contained 155 unique compounds with reported IC_{50} values. ECFP4 fingerprints were generated and Tanimoto similarity coefficients calculated between each candidate compound and every compound in both reference sets.

ZINC9412756 showed a maximum ECFP4 Tanimoto similarity of 0.3827 to its closest known MDM2 inhibitor (ChEMBL3691703) and 0.2500 to its closest MDM4 inhibitor (ChEMBL5199785) (**Fig. 4**). ZINC20724474 showed even lower similarities: 0.3488 to the closest MDM2 inhibitor (ChEMBL3691703) and 0.2315 to the closest MDM4 inhibitor (ChEMBL3601398). Both values fall well below the 0.5 threshold commonly used to define structural relatedness, indicating that both compounds occupy regions of chemical space largely unexplored by existing MDM2/MDM4 inhibitors. By comparison, the maximum similarity between any two known MDM2 inhibitors in the ChEMBL set averages 0.65–0.80, highlighting that our candidate compounds are substantially more divergent from known chemical space than typical diversity among established inhibitors. Novelty scores of 0.6173–0.7685 further quantify the distance of these candidate compounds from known chemical space.

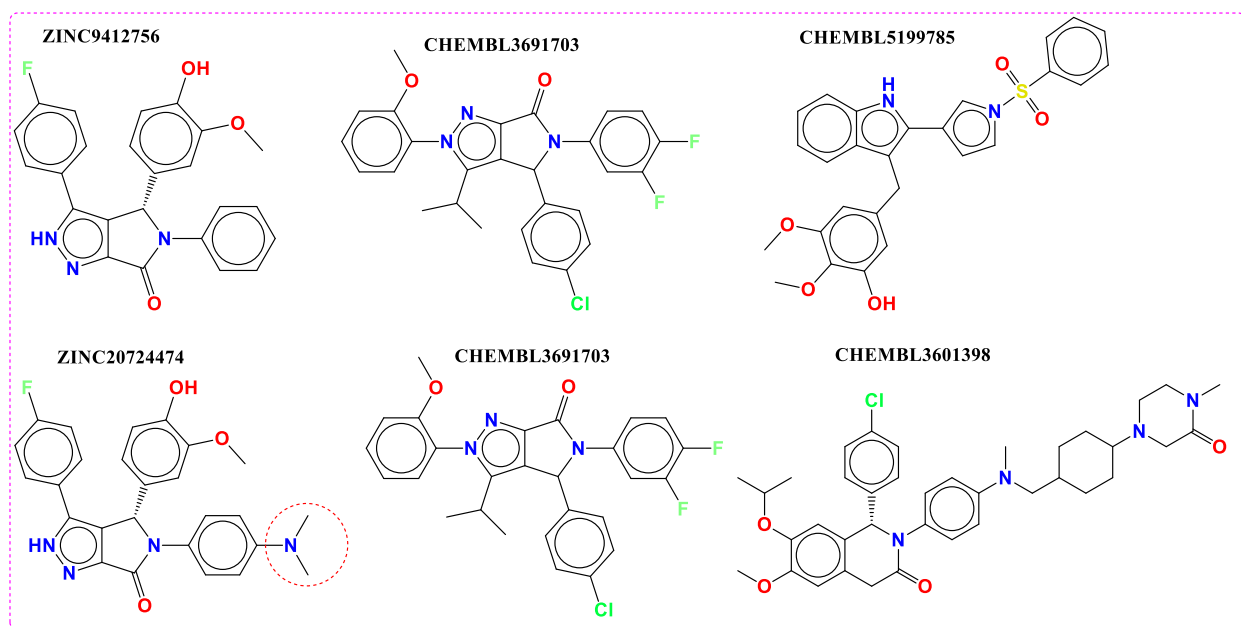


Fig. 4: Chemical structures of the identified candidate compounds, ZINC9412756 and ZINC20724474, with their structural novelty assessment against known MDM2 and MDM4 inhibitors from the ChEMBL database.

3.3. Molecular Dynamics Simulations

To probe the dynamic behavior of ligand binding, 500-ns all-atom simulations were performed for apo MDM2, apo MDM4, and their complexes with ZINC9412756 and ZINC20724474 (**Figs. 5-7**).

3.3.1. Binding stability and interaction profiles for MDM2 complexes

Both compounds remained stably bound to MDM2 throughout the simulation, with low and narrowly distributed ligand RMSD values (ZINC9412756: 0.099 ± 0.021 nm; ZINC20724474: 0.127 ± 0.022 nm). Protein-ligand RMSD plateaued within the first 50 ns (0.681 ± 0.065 nm and 0.670 ± 0.100 nm, respectively). Visual inspection of trajectories confirmed that neither ligand dissociated, as evidenced by minimal fluctuations after full equilibration (>50 ns) and narrow standard deviations. Interaction analysis revealed binding modes closely resembling the static docking poses, with stability conferred by dense, persistent networks of non-bonded contacts and hydrogen bonds. In the ZINC20724474-bound complex, pocket residues Leu54, Phe55, Gly58, Gln59, Val93, and His96 remained in contact with the ligand for >50% of the simulation; persistent hydrogen bonds were observed with Gln59 (67.5%), His96 (61.9%), and Val93 (55.9%). The ZINC9412756 complex displayed an equally robust contact map involving the same residues, all with >91% persistence, together with hydrogen bonds to Val93 (68.3%), His96 (59.9%), Gly58 (52.4%), and Gln59 (43.2%). These persistent interactions likely act as molecular tethers, accounting for the stability and small protein-ligand RMSD deviations observed.

Despite comparable binding stability, the two ligands reshaped MDM2 dynamics in slightly different ways. Backbone RMSD was elevated and more variable for ZINC9412756 (0.230 ± 0.102 nm) relative to ZINC20724474 (0.165 ± 0.032 nm) and apo (0.137 ± 0.019 nm), while the radius of gyration increased modestly only with ZINC9412756 (1.336 ± 0.028 nm vs. apo 1.314 ± 0.010 nm).

Per-residue RMSF analysis within the p53-binding domain showed that ZINC9412756 reduced fluctuations of immediate contact residues (Leu54, Leu57, Gly58, Ile61, Met62; 6–10% reduction) but markedly increased mobility in the Val93-Phe91 loop (Phe91: +48%, Val93: +58%) and flanking hydrophobic floor residues (His96: +26%, Ile99: +40%, Tyr100: +44%). By contrast, ZINC20724474 dampened fluctuations across most of the p53-binding-domain core (Lys51, Leu54, Phe55, Leu57, Gly58, Gln59, Ile61, Met62, His96; 6–12% reduction) and only modestly enhanced motions at peripheral sites (Phe86, Phe91, Val93, Ile99; 2–15%). These contrasting mobility patterns appear to originate from differences in hydrogen-bond topology: ZINC20724474 forms a distributed polar network engaging Gln59, His96, and Val93, supplemented by contacts with Lys51, Leu54, Gly58, Ile61, and Tyr67, whereas ZINC9412756 concentrates its hydrogen bonds onto a few central anchors (Val93, His96, Gly58), leaving flanking regions less restrained.

Principal component analysis (PCA) of C α motions reinforced this observation. Apo MDM2 distributed variance over PC1 (75.9%) and PC2 (24.1%). ZINC20724474 broadened the PC2 contribution (PC1 61.5%, PC2 38.5%) and modestly expanded sampled ranges (PC1 5.05 nm, PC2 3.95 nm). ZINC9412756,

however, concentrated 86.2% of variance into PC1 and extended its range to 7.47 nm, indicating motion dominated by a single collective coordinate. The free-energy landscape constructed from these PCs contained four minima for apo MDM2 with a barrier of 2.09 kT and a heterogeneity index of 16.9. ZINC20724474 produced five minima with comparable barrier (2.09 kT) and index of 19.9, whereas ZINC9412756 generated ten minima with a barrier of 7.35 kT and index of 43.1, consistent with a rugged landscape of slowly interconverting substates.

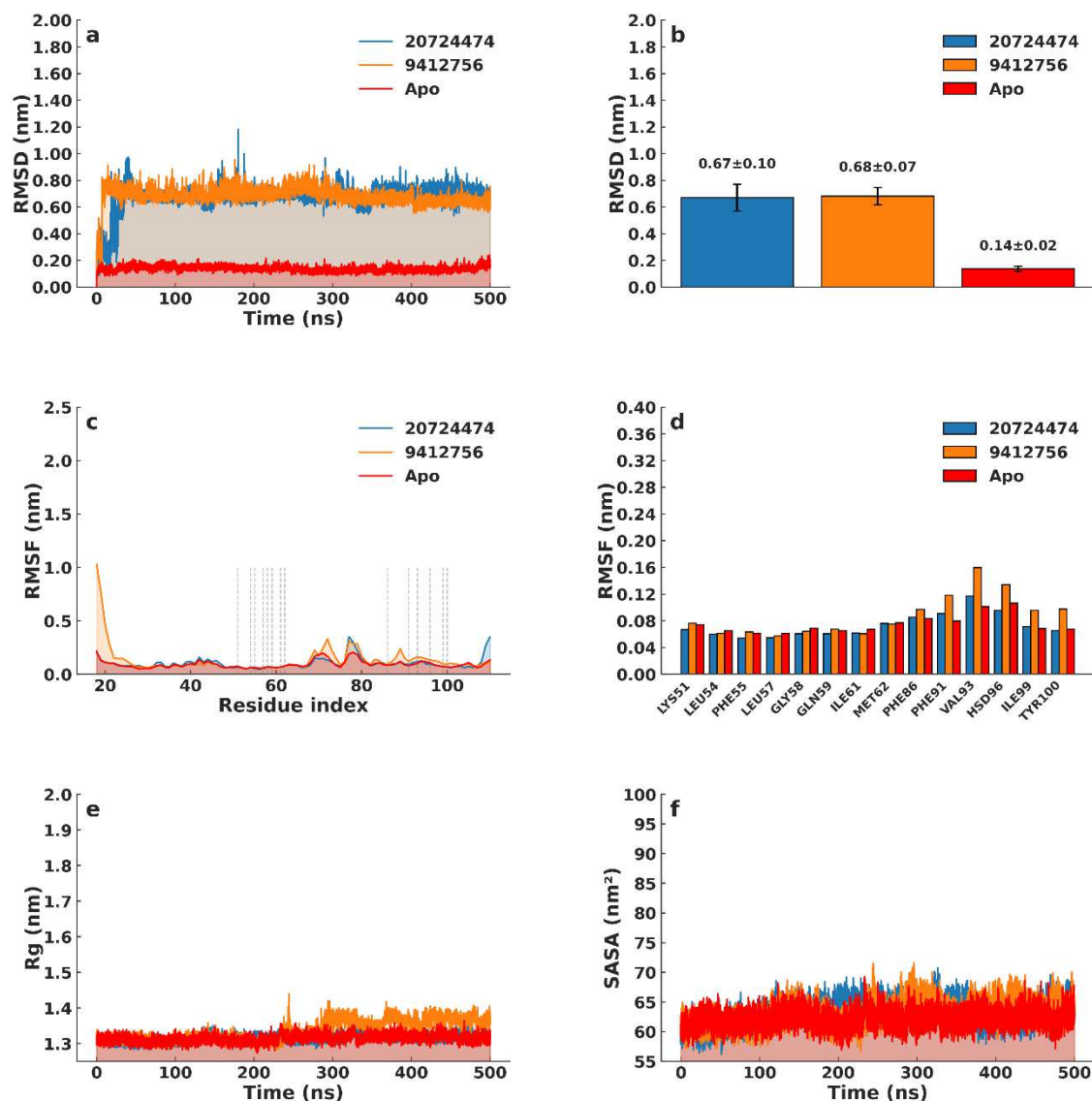


Fig. 5: Structural stability of MDM2 complexes during 500 ns molecular dynamics simulations. **a**, Root-mean-square deviation (RMSD) of protein α atoms and ligand heavy atoms as a function of time. **b**, Mean RMSD $\pm$ s.d. for each bound complex and for apo MDM2. **c**, Per-residue root-mean-square fluctuation (RMSF) of α atoms. **d**, α RMSF of key active-site residues. **e**, Radius of gyration (R_g). **f**, Solvent-accessible surface area (SASA).

3.3.2 Binding Stability and Interaction Profiles for MDM4-complexes

For the MDM4 homologue, both ligands also remained stably bound. Ligand RMSD values were 0.122 ± 0.019 nm (ZINC9412756) and 0.144 ± 0.049 nm (ZINC20724474), with protein-ligand RMSD converging after approximately 250 ns to 0.599 ± 0.132 nm and 0.552 ± 0.096 nm, respectively. Visual inspection of trajectories confirmed that neither ligand dissociated from MDM4 throughout the simulation. Interaction analysis revealed that in the MDM4-ZINC20724474 complex, residues Met52, Leu55, Gly56, Ile59, Pro94, Phe89, Val91, and Leu97 remained in contact with the ligand for >95% of the simulation. The polar network was dominated by two persistent hydrogen bonds to Val91 (89.3%) and Met52 (82.8%), supplemented by weaker interactions with Lys92, Ile59, Gly56, and Leu97. ZINC9412756 maintained an equally extensive contact network (>95% persistence for the same residues) but exhibited a less extensive hydrogen-bond network, with hydrogen bonds to Val91 (62.9%) and Met52 (69.9%) complemented by weaker contacts with His53 (39.9%), Gly56 (42.2%), and Leu97 (31.8%). These interaction profiles are consistent with docking-predicted protein-ligand interactions, supporting the docked models.

RMSF analysis of the MDM4 complexes revealed that ZINC20724474 increased fluctuations in the C-terminal portion of the p53-binding domain (Leu97: +47%, Tyr98: +31%, Leu101: +28%, Pro94: +37%) and in more distal elements (Gln70: +13%, Phe89: +23%), while slightly reducing fluctuations in the central core (Gly56, Ile59, Met60, Val91; 5–10% decrease). ZINC9412756, by contrast, rigidified the same C-terminal region (Leu97: -13%, Tyr98: -3.6%, Leu101: -24%) and increased mobility in most other pocket residues (Met52: +23%, Leu55: +19%, Gly56: +7.6%, Ile59: +11%, Met60: +10%, Gln70: +21%, Val91: +7.2%). These divergent mobility patterns coincide with differences in hydrogen-bond topology. In the ZINC20724474 complex, persistent hydrogen bonds with Val91 and Met52 appear to restrict the ligand at the centre and N-terminal side of the p53-binding domain. This restriction is accompanied by enhanced fluctuations at the C-terminal end. In contrast, ZINC9412756 exhibits a more diffuse polar network, correlating with stabilization of the C-terminal region and increased mobility in the central pocket.

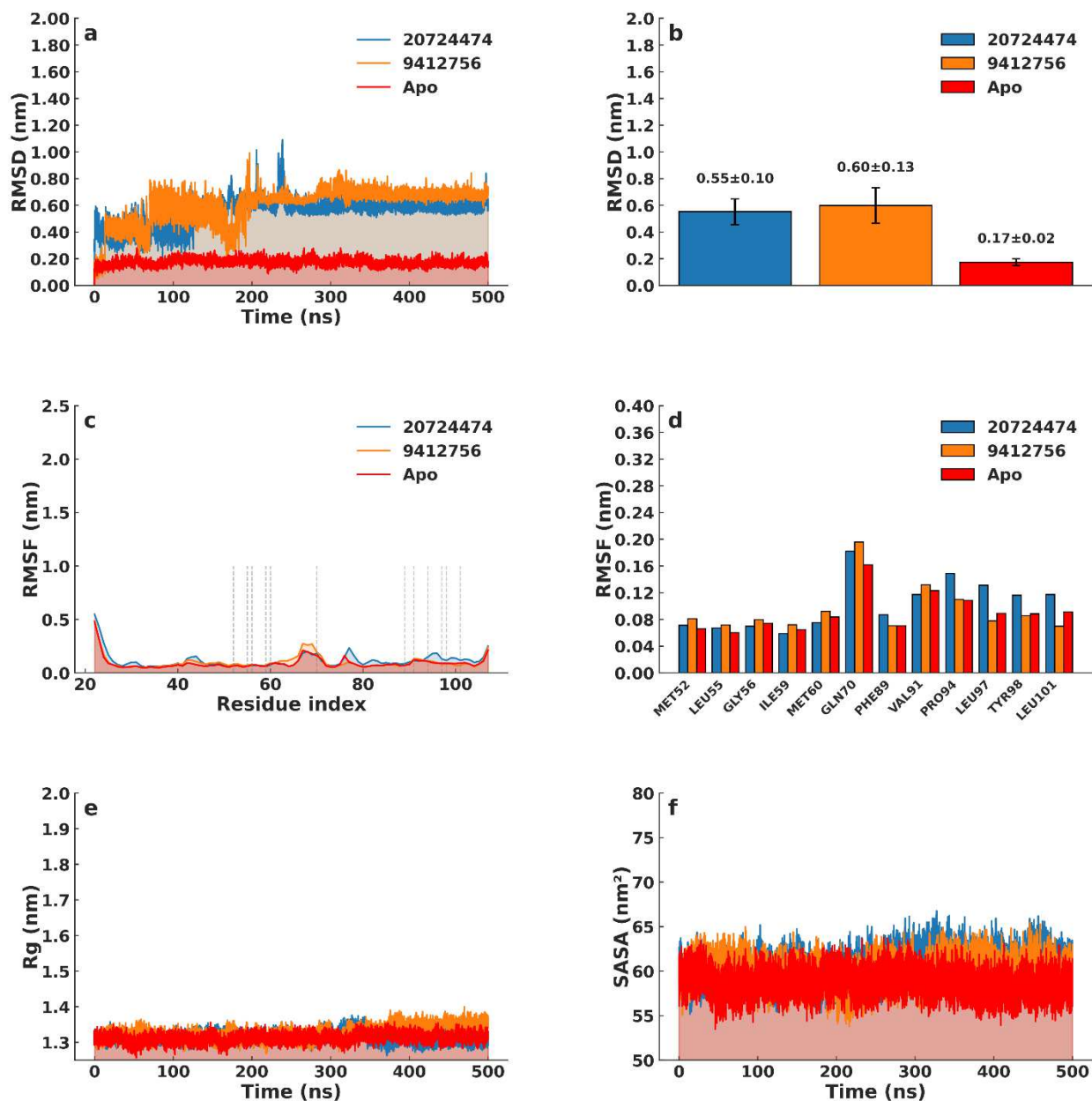


Fig. 6: Structural stability of MDM4 complexes during 500 ns molecular dynamics simulations. **a**, Root-mean-square deviation (RMSD) of protein C α atoms and ligand heavy atoms as a function of time. **b**, Mean RMSD $\pm$ s.d. for each bound complex and for apo MDM2. **c**, Per-residue root-mean-square fluctuation (RMSF) of C α atoms. **d**, C α RMSF of key active-site residues. **e**, Radius of gyration (R_g). **f**, Solvent-accessible surface area (SASA).

Principal component analysis (PCA) of MDM4 trajectories showed that ligand binding redistributed variance towards PC1. In the apo system, PC1 and PC2 accounted for 54.9% and 45.1% of total variance, respectively. For ZINC20724474, contributions shifted to PC1: 66.7%, PC2: 33.3%; for ZINC9412756, PC1: 65.9%, PC2: 34.1%. Conformational space along PC1 expanded to 7.30 nm (PC2: 4.51 nm) in the

ZINC20724474 complex, whereas ZINC9412756 exhibited a smaller increase (5.89 nm; PC2: 4.73 nm). The free-energy landscape of apo MDM4 contained four distinct minima with a barrier of 2.30 kT and complexity index of 19.8. ZINC9412756 preserved four minima but lowered the barrier to 1.79 kT and raised the complexity index to 27.8. ZINC20724474 increased minima to six, maintained a barrier of 2.05 kT, and produced a complexity index of 32.8, indicating a more rugged conformational ensemble.

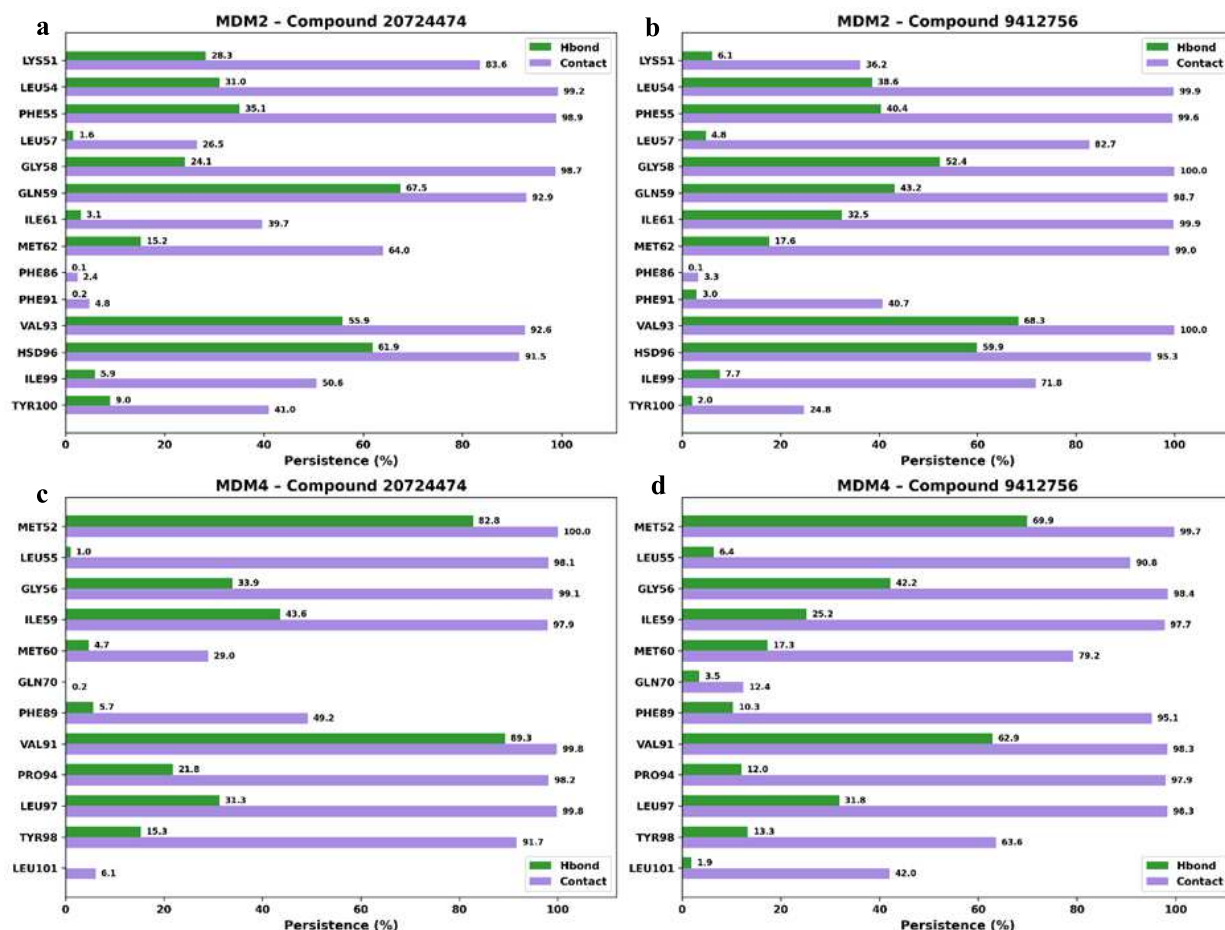


Fig. 7: Residue-wise persistence, hydrogen bonding, and contact frequencies for two hit compounds. (a) Hydrogen bond persistence frequencies for ZINC20724474-MDM2. **(b)** Hydrogen bond persistence frequencies for ZINC9412756-MDM2. **(c)** Hydrogen bond persistence frequencies for ZINC20724474-MDM4. **(d)** Hydrogen bond persistence frequencies for ZINC9412756-MDM4

3.4.0 MM/GBSA Binding Free Energies of the MDM2 and MDM4 Complexes

MM/GBSA calculations were used to estimate the binding free energies of ZINC9412756 and ZINC20724474 with MDM2 and MDM4, complementing the static docking predictions. Across all four simulated systems, the computed ΔG_{bind} values were consistently negative (**Table 2, Figs. S3, S4**): MDM2-ZINC9412756, -23.98 ± 1.69 kcal/mol; MDM2-ZINC20724474, -20.73 ± 4.22 kcal/mol;

MDM4-ZINC9412756, -24.43 ± 4.17 kcal/mol; and MDM4-ZINC20724474, -26.60 ± 3.68 kcal/mol. These results are consistent with the docking-derived hypothesis that both compounds may act as dual MDM2/MDM4 inhibitors. ZINC9412756 bound both targets with comparable computed affinities ($\Delta\Delta G \approx 0.45$ kcal/mol), whereas ZINC20724474 exhibited a 5.87 kcal/mol preference for MDM4 over MDM2. This differential profile may have pharmacological implications: a balanced inhibitor such as ZINC9412756 could simultaneously engage both MDM2 and MDM4, while an MDM4-preferring inhibitor such as ZINC20724474 might preferentially target tumors that overexpress MDM4.

The gas-phase contribution is the primary driver of binding thermodynamics, with van der Waals (ΔE_{VDW}) and electrostatic (ΔE_{ELE}) components providing the main stabilizing forces. For MDM2 complexes, ΔE_{VDW} and ΔE_{ELE} were -36.37 ± 0.97 and -8.06 ± 0.31 kcal/mol (ZINC9412756), and -32.42 ± 0.53 and -12.67 ± 3.91 kcal/mol (ZINC20724474). For MDM4, corresponding values were -35.76 ± 1.56 and -5.12 ± 3.56 kcal/mol (ZINC9412756), and -34.89 ± 1.08 and -15.52 ± 3.19 kcal/mol (ZINC20724474). Van der Waals interactions accounted for approximately 70-85% of gas-phase stabilization across all systems, consistent with the predominantly hydrophobic character of the p53-binding clefts. The electrostatic component, although smaller, is consistent with hydrogen bonding interactions identified in docking poses and MD trajectories. The solvation penalty partially offsets gas-phase stabilization but does not negate the net favourable binding free energy.

The values of -20 to -27 kcal/mol obtained here suggest binding to MDM2/MDM4. However, MM/GBSA is known to systematically overestimate affinities, particularly for lipophilic ligands, and the absolute energies are not directly comparable to experimental values. Accordingly, these results were interpreted solely as a comparative ranking, and experimental validation will be required to establish true binding affinities.

Table 2. Mean binding free energies and component contributions ($\pm$ s.d.) for ZINC9412756 and ZINC20724474 binding to MDM2 and MDM4.

	Total	ΔG_{SOLV}	ΔG_{GAS}	ΔE_{SURF}	ΔE_{GB}	ΔE_{EL}	$\Delta E_{VDWAALS}$
MDM2							
ZINC9412756	-23.98 ± 2.1	20.44 ± 4.53	-44.42 ± 5.7	-4.85 ± 0.33	25.29 ± 4.6	-8.06 ± 5.1	-36.37 ± 2.94
ZINC20724474	-20.73 ± 4.1	24.36 ± 4.58	-45.09 ± 7.1	-4.14 ± 0.49	28.50 ± 4.8	-12.67 ± 5	-32.42 ± 4.17
MDM4							
ZINC9412756	-24.43 ± 3.1	16.45 ± 3.54	-40.88 ± 5.6	-4.73 ± 0.53	21.18 ± 3.6	-5.12 ± 5.1	-35.76 ± 3.97
ZINC20724474	-26.60 ± 2.1	23.81 ± 4.29	-50.42 ± 5.4	-4.57 ± 0.33	28.38 ± 4.4	$-15.502 \pm$	-34.89 ± 3.14

Per-residue energy decomposition identified key hotspot residues stabilizing the complexes. In the MDM2 pocket, van der Waals and non-polar solvation terms dominate. For ZINC20724474, the largest total contribution arises from Phe55 ($\Delta G_{\text{total}} = -2.62 \pm 0.05$ kcal/mol), followed by Leu54 (-1.54 ± 0.05), Met62 (-1.44 ± 0.07), Gly58 (-1.23 ± 0.03), and Gln59 (-1.00 ± 0.07). The van der Waals component at Phe55 (-2.88 ± 0.02 kcal/mol) reflects dense hydrophobic packing with the ligand's aromatic ring. Gln59 exhibits a larger electrostatic contribution (-1.43 ± 0.05) partly offset by polar solvation ($+1.28 \pm 0.02$), consistent with the persistent hydrogen bond observed in MD trajectories. ZINC9412756 shows a qualitatively similar profile with systematically smaller total contributions: Phe55 (-1.63 ± 0.05), Leu54 (-1.68 ± 0.06), Gly58 (-1.62 ± 0.03), and Met62 (-1.32 ± 0.06). The van der Waals term at Phe55 for ZINC9412756 is -1.86 ± 0.01 kcal/mol, approximately 29% weaker than for ZINC20724474. This reduced hydrophobic anchor may account for the higher mobility of the adjacent Val93-Phe91 loop and the more rugged free-energy landscape observed for this complex. For both compounds, >90% of stabilization at Leu54, Phe55, Met62, and Gln59 originates from side-chain atoms, confirming that interactions mimic side-chain insertions of p53.

On MDM4, the energetic profiles differ between ligands. For ZINC20724474, Met52 provides the largest stabilization ($\Delta G_{\text{total}} = -3.60 \pm 0.05$ kcal/mol), arising from a large van der Waals term (-3.39 ± 0.01) and substantial electrostatic component (-1.93 ± 0.02) partially compensated by polar solvation ($+2.05 \pm 0.01$), consistent with the persistent hydrogen bond to the Met52 backbone. Pro94 contributes -2.78 ± 0.05 kcal/mol (van der Waals -2.72 , electrostatic -0.74), and Val91 adds -1.61 ± 0.05 (electrostatic -2.76 , van der Waals -0.67). The strong electrostatic signature at Val91 corresponds to the hydrogen bond present for 89.3% of the trajectory. For ZINC9412756 on MDM4, total contributions are generally smaller: Met52 (-2.39 ± 0.06), Pro94 (-1.25 ± 0.04), Gly56 (-1.11 ± 0.03), Leu97 (-1.29 ± 0.05), and Val91 (-1.00 ± 0.10). The electrostatic terms at Met52 (-0.42 ± 0.03) and Val91 (-0.23 ± 0.07) are modest, consistent with the absence of persistent conventional hydrogen bonds to these residues in the MDM4-ZINC9412756 trajectory.

The key hotspot residues identified (Phe55, Leu54, Met62, and Gln59 in MDM2; Met52, Pro94, and Val91 in MDM4) correspond to established anchoring points for p53-derived peptides and small-molecule inhibitors. Side-chain interactions account for >80% of total stabilization at these positions, indicating that both compounds act as structural mimics of the p53 triad (Phe19, Trp23, Leu26).

3.5.0 Network pharmacology

Differential expression analysis of the GSE21122 dataset (46 DDLPS tumours versus 9 normal adipose controls; $\text{adj.P.Val} < 0.05$, $|\log\text{FC}| \geq 1$) identified 1,354 genes. Intersection with 517 DDLPS-associated genes from GeneCards yielded 101 high-confidence disease genes supported by both transcriptomic and literature evidence. Independently, SwissTargetPrediction (probability ≥ 0.1) returned 104 non-redundant putative targets for ZINC20724474 and ZINC9412756. The overlap of the compound-predicted target set with the 101 DDLPS-enriched genes comprised ten proteins: MDM2, MDM4, EGFR, IDH1, PIM1, AURKA, FLT1, MME, CDK1, and CDK4. This intersection defines the DDLPS-relevant polypharmacology predicted for the two candidate compounds.

The ten targets formed a connected protein-protein interaction (PPI) network with 10 nodes and 19 edges, an average node degree of 3.8, a local clustering coefficient of 0.66, and a PPI enrichment p-value of 6.82×10^{-5} . This strong enrichment indicates that the observed interactions reflect genuine functional associations. The network recapitulates the hallmark co-amplification of *MDM2* and *CDK4* on chromosome 12q13-15 and the characterized MDM2-MDM4 heterodimerisation that regulates p53 degradation. Cell-cycle regulators (CDK1, AURKA, PIM1) and receptor tyrosine kinases (EGFR, FLT1) complete the module.

Functional enrichment was performed using STRING (**Fig. S5**). The most significantly enriched KEGG pathway was p53 signalling (hsa04115; observed 4 genes, strength 2.04, $\text{FDR} = 1.42 \times 10^{-5}$), involving CDK4, MDM2, MDM4, and CDK1. This aligns with the intended dual MDM2/MDM4 inhibitory mechanism and implicates additional kinases that phosphorylate p53 or MDM2. Other enriched KEGG pathways included *MicroRNAs in cancer* ($\text{FDR} = 1.5 \times 10^{-4}$), *Bladder cancer* ($\text{FDR} = 1.5 \times 10^{-4}$), *Glioma* ($\text{FDR} = 5 \times 10^{-4}$), *Melanoma* ($\text{FDR} = 5 \times 10^{-4}$), *PI3K-Akt signalling* ($\text{FDR} = 9.4 \times 10^{-4}$), and *Cell cycle* ($\text{FDR} = 1.2 \times 10^{-3}$). The presence of the PI3K-Akt pathway is noteworthy because Akt-mediated phosphorylation stabilises MDM2 and promotes its nuclear import; concurrent inhibition of EGFR and FLT1, both upstream activators of PI3K-Akt, could therefore enhance p53 reactivation beyond direct MDM2/MDM4 blockade.

Gene Ontology biological process analysis reinforced the cell-cycle-centric nature of the module. The top term was *positive regulation of cell cycle* (GO:0045787; 6 genes, $\text{FDR} = 1.0 \times 10^{-4}$), followed by *positive regulation of cell cycle process* (GO:0090068), *positive regulation of mitotic cell cycle phase transition* (GO:1901992), and *regulation of cell cycle* (GO:0051726). These annotations reflect the roles of CDK4, CDK1, AURKA, and PIM1 in G1/S and G2/M transitions, and of MDM2 in suppressing p53-dependent cell-cycle arrest. *Negative regulation of apoptotic process* (GO:0043066; 6 genes, $\text{FDR} = 2.2 \times 10^{-3}$) involved MDM2, MDM4, EGFR, PIM1, AURKA, and CDK1. Additional terms linked to protein stability, including *regulation of protein stability* (GO:0031647; 4 genes, $\text{FDR} = 9.2 \times 10^{-3}$) and *regulation*

of *protein catabolic process* (GO:0042176), are consistent with the E3 ubiquitin ligase activity of MDM2 and its modulation by MDM4.

Among the most specific disease terms from the DISEASES database was *dedifferentiated liposarcoma* (DOID:0080531; CDK4 and MDM2, strength 3.12, FDR = 1.3×10^{-3}). Other sarcoma subtypes, including bone sarcoma, pleomorphic liposarcoma, and well-differentiated liposarcoma, highlight the relevance of the target set across the liposarcoma spectrum. IDH1, although rarely mutated in DDLPS, has been linked to metabolic reprogramming and chemoresistance in soft-tissue sarcomas when overexpressed. These analyses indicate that ZINC20724474 and ZINC9412756 are predicted to engage a DDLPS-relevant target module that converges on p53 regulation and cell-cycle control. The physical and functional connections among these targets, exemplified by the direct MDM2-CDK4 and MDM2-MDM4 interactions, suggest that simultaneous engagement of multiple network hubs might produce cooperative effects, potentially addressing the limited activity of selective MDM2 inhibitors seen in clinical trials. The enrichment of the p53 pathway, cell cycle, and PI3K-Akt signalling provides a mechanistic rationale for exploring combination strategies. The presence of CDK4, a genetically co-amplified driver, raises the possibility of CDK4 engagement; however, experimental validation is required to determine whether the predicted target interactions translate into functional inhibition. The network topology, with its high clustering coefficient and significant PPI enrichment, supports the concept that polypharmacology within this interconnected module represents a distinct approach to targeting DDLPS compared with single-agent MDM2 inhibition.

3.6. ADME profiling

Physicochemical properties, drug-likeness, absorption, distribution, metabolism, excretion, and toxicity were predicted for ZINC9412756, ZINC20724474, and nutlin-3a (as a clinical reference) using SwissADME and ADMETlab 3.0 (**Table S7**).

All three compounds satisfy Lipinski's rule of five and the Pfizer toxicity rule, indicating acceptable oral drug-likeness and low inherent toxicity risk by these criteria. They violate the GSK rule ($MW > 400$) and, in the case of nutlin-3a, the Golden Triangle ($MW > 500$), reflecting their large molecular size. Quantitative estimates of drug-likeness (QED) are moderate (ZINC20724474, 0.451; ZINC9412756, 0.507; nutlin-3a, 0.394) and lie below the mean of attractive compounds (0.67). The fraction of sp^3 -hybridised carbons (F_{sp^3}) is low (0.15, 0.08, 0.30), consistent with the highly aromatic character of these scaffolds; however, MCE-18 scores are high, indicating structurally distinctive chemotypes. Natural-product-likeness scores are negative (−0.82, −0.67, −0.25), confirming their synthetic origin. No pan-assay interference compounds (PAINS) alerts were identified, and all three compounds are free of BMS reactive alerts. However, structural alerts for thiol reactivity (Alarm NMR) are present in all three, and the ZINC compounds additionally contain chelating motifs. Synthetic accessibility is favourable, with SAScore values and GASA probabilities predicting that the compounds are easy to prepare.

Caco-2 permeability values exceed the favourable threshold of $-5.15 \log \text{ cm s}^{-1}$ for ZINC9412756 (−5.14) and nutlin-3a (−4.92), while ZINC20724474 falls marginally short (−5.19). PAMPA yields low probabilities of high permeability for all compounds (0.025, 0.032, 0.00039), suggesting passive diffusion alone is not the primary absorption route. Aqueous solubility ($\log S$) is modest (−5.44, −5.22, and −4.49, respectively), indicating solubility-limited dissolution could constrain oral absorption. The dimethylamino substituent in ZINC20724474 increases P-gp inhibitory potential (probability 0.996) but also raises the probability of being a P-gp substrate (0.909 vs. 0.656 for ZINC9412756 and 0.00073 for nutlin-3a), which may reduce intracellular accumulation. All three compounds are predicted to inhibit P-gp with high probability (≥ 0.989). Predicted human intestinal absorption is excellent, with extremely low probabilities of poor absorption. The probability of oral bioavailability falling below 20% ($F_{20\%+}$) is low (0.032, 0.015, 0.00021), supporting favourable systemic exposure, though high oral bioavailability ($F_{50\%+}$) is unlikely without formulation or structural optimisation (0.84, 0.75, and 0.30, respectively).

Distribution properties are primarily governed by lipophilicity, plasma protein binding, and transporter interactions. All three compounds exhibit high predicted plasma protein binding exceeding 97% (ZINC20724474, 97.81%; ZINC9412756, 98.39%; nutlin-3a, 99.19%), with correspondingly low unbound fractions (1.94%, 1.42%, and 0.73%). Predicted steady-state volumes of distribution ($\log V_{Dss}$: 0.245, 0.157, and 0.355 L/kg) are modest, indicating limited tissue distribution. Blood-brain barrier penetration differs markedly: both ZINC20724474 and ZINC9412756 are predicted to be BBB-negative (probability of

BBB+ 0.0016 and 0.151, respectively), consistent with their strong P-gp substrate behaviour. Nutlin-3a, with negligible P-gp substrate probability, shows a moderate probability of BBB penetration (0.339).

CYP inhibition and substrate probabilities were predicted for major hepatic isoforms. All three compounds show substantial inhibitory potential, particularly against CYP2C9 (probabilities of 1.000 for the ZINC compounds and 0.811 for nutlin-3a) and CYP2C19 (0.976 for ZINC9412756, 0.989 for nutlin-3a). CYP3A4 inhibition probabilities are 0.671, 0.964, and 0.126 for ZINC20724474, ZINC9412756, and nutlin-3a, respectively. CYP2C8 inhibition is universally high (probabilities ~1.0). CYP3A4 substrate probability is high for ZINC20724474 (0.984) and ZINC9412756 (0.999), but notably low for nutlin-3a (0.00497), suggesting the clinical compound largely avoids CYP3A4-mediated clearance. Human liver microsomal stability predictions indicate ZINC20724474 has a low probability of instability (0.0773), ZINC9412756 intermediate (0.214), and nutlin-3a high (0.964), predicting rapid microsomal degradation for the reference compound.

Plasma clearance (CL_{plasma}) is low for all compounds (3.556, 2.806, and 3.785 mL min⁻¹ kg⁻¹), consistent with extensive protein binding. However, predicted half-life (t_{1/2}) values are short: 0.763 h (ZINC20724474), 0.861 h (ZINC9412756), and 1.031 h (nutlin-3a), falling in the ultra-short to short categories. The combination of low clearance and short half-life is consistent with modest predicted volumes of distribution, suggesting frequent dosing could be required to maintain therapeutic concentrations.

Toxicity predictions were obtained for a panel of endpoints covering cardiac, hepatic, renal, genetic, and other organ toxicities, as well as nuclear receptor and stress response pathway modulation. The probability of hERG blockade (IC₅₀ ≤ 10 μM) was highest for nutlin-3a (0.767), intermediate for ZINC9412756 (0.195), and lowest for ZINC20724474 (0.172). Hepatotoxicity and nephrotoxicity liabilities were consistently high for all three compounds, with drug-induced liver injury (DILI) probabilities exceeding 0.94 and nephrotoxicity probabilities exceeding 0.98. Genetic toxicity endpoints revealed uniformly high genotoxicity probabilities (≥0.987), while carcinogenicity probability was moderate for ZINC20724474 (0.612) and nutlin-3a (0.376) and lower for ZINC9412756 (0.272). Ames mutagenicity probability was 0.706 (ZINC20724474), 0.429 (ZINC9412756), and 0.0891 (nutlin-3a). Acute oral toxicity was moderate for the ZINC compounds (0.634, 0.501) and low for nutlin-3a (0.264). Skin sensitisation probability was high (0.792, 0.748, and 0.783), and respiratory toxicity probability was high for ZINC20724474 (0.986), moderate for ZINC9412756 (0.790), and low for nutlin-3a (0.201). Nuclear receptor modulation varied: AhR activation was high for ZINC20724474 (0.996) and ZINC9412756 (0.880) but low for nutlin-3a (0.0422); ER agonism was high for ZINC20724474 (0.892), moderate for ZINC9412756 (0.561), and low for nutlin-3a (0.00653). Stress response pathway activation indicated potential mitochondrial toxicity (MMP disruption) for both ZINC compounds (0.995, 0.973) but not nutlin-3a (0.0133). p53 pathway activation followed a similar pattern (0.837, 0.303, and 0.00196), consistent with the MDM2/MDM4 inhibitory activity predicted for the ZINC compounds.

4.0. Discussions

In this study, we employed an integrated computational pipeline combining ensemble machine-learning QSAR, ensemble docking across 23 MDM2 and 24 MDM4 crystal structures, molecular dynamics simulations, MM/GBSA free-energy calculations, network pharmacology, and ADMET profiling to prioritize candidate dual MDM2/MDM4 inhibitors for DDLPS. Virtual screening of 9.4 million drug-like compounds identified ZINC9412756 and ZINC20724474 as candidate compounds with predicted affinity for both MDM2 and MDM4. The Random Forest QSAR model achieved test $R^2 = 0.76$, RMSE = 0.64, and CCC = 0.87, with y-randomization indicating genuine structure-activity relationships. Ensemble docking predicted binding affinities of -9.10 ± 0.30 to -9.27 ± 0.30 kcal/mol for both targets, exceeding those of co-crystallised reference inhibitors. The binding modes are characterized by extensive hydrophobic contacts, with van der Waals interactions contributing 70-90% of gas-phase stabilization, and persistent hydrogen bonds to key residues including Tyr100 (MDM2) and Met52 (MDM4). MM/GBSA calculations yielded binding free energies of -20.73 ± 4.02 to -26.60 ± 2.76 kcal/mol; however, MM/GBSA is known to systematically overestimate affinities for lipophilic ligands, and these values should be interpreted as comparative rankings rather than quantitative predictions. Network pharmacology predicted engagement of a DDLPS-relevant module (MDM2, MDM4, CDK4, EGFR, AURKA) enriched in p53 signaling ($FDR = 1.42 \times 10^{-5}$). ADMET predictions indicated favorable oral bioavailability and synthetic accessibility, with several toxicity alerts (hepatotoxicity, nephrotoxicity, hERG liability) that represent tractable optimization targets. Structural similarity analysis against known MDM2 ($n = 4,154$) and MDM4 ($n = 155$) inhibitors indicated that both candidates are structurally distinct from previously reported chemotypes, with maximum ECFP4 Tanimoto similarities of 0.3827 and 0.3488 to the closest MDM2 inhibitors and 0.2500 and 0.2315 to the closest MDM4 inhibitors, values far below the 0.5 threshold for relatedness. This suggests that these compounds occupy chemical space largely unexplored by existing MDM2/MDM4 inhibitors and may offer distinct binding modes.

Previous virtual screening efforts for MDM2/MDM4 inhibitors have typically relied on single-docking protocols and smaller compound libraries. The discovery of YL93 involved structure-based design starting from known MDM2 scaffolds [36], while other dual inhibitors have been identified through fragment-based approaches or repurposing screens [32]. However, our integrated pipeline enabled screening of a substantially larger chemical space (9.4 million compounds) and incorporated multiple receptor conformations to account for protein flexibility. The ensemble docking approach, using 23 MDM2 and 24 MDM4 structures, is more comprehensive than typical single-structure docking and may better identify binders that accommodate conformational variability. The combination of ML-QSAR filtering prior to docking also represents a more efficient screening strategy than physics-based docking alone, reducing the number of candidates requiring computationally expensive evaluation from 9.4 million to 89 compounds. The identified candidate compounds are structurally distinct from known MDM2 inhibitors such as nutlin-3a, RG7112, and AMG-232. While nutlin-class inhibitors rely on cis-imidazoline cores with halogenated phenyl

groups, ZINC9412756 and ZINC20724474 feature pyrazole and phenolic scaffolds. The binding mode analysis reveals a hydrogen bond to Tyr100 in MDM2 that is not observed in the co-crystallised controls, suggesting a potentially different mode of p53-mimicry. In MDM4, ZINC9412756 forms an extensive polar interaction network with three conventional hydrogen bonds to Met52, His53, and Gln70, which is not observed in any co-crystallised MDM4 inhibitor evaluated. If validated experimentally, these candidate compounds could address the unmet need for more effective MDM2/MDM4-targeted therapies in DDLPS. The dual-targeting approach is particularly relevant given the limited clinical activity of MDM2-selective inhibitors and the emerging evidence that MDM4 upregulation mediates resistance [24,25,34]. Clinical experience with RG7112, MI-77301, and siremadlin has demonstrated that while MDM2 inhibition can achieve disease stabilization in some DDLPS patients, durable objective responses are rare and dose-limiting toxicities are common [23–25]. A strategy that simultaneously targets both MDM2 and MDM4 may achieve more complete p53 reactivation and overcome MDM4-mediated resistance.

The network pharmacology predictions suggest that these compounds may engage additional DDLPS-relevant targets (CDK4, EGFR, AURKA), potentially offering polypharmacological benefits. Simultaneous inhibition of MDM2, MDM4, and CDK4 is notable given their co-amplification in DDLPS [8,9], and EGFR/AURKA inhibition could address other oncogenic signaling pathways. This polypharmacological profile may enhance efficacy compared to single-target inhibition, though it also raises the potential for increased off-target toxicity that would require careful evaluation during lead optimization.

The predicted ADMET profile of the candidate compounds is generally favorable for further development. ZINC9412756 emerges as the more balanced candidate based on its lower hERG liability and P-gp substrate potential. However, the toxicity alerts for hepatotoxicity, nephrotoxicity, and hERG liability will require careful experimental evaluation and structure-based optimization. The synthetic accessibility scores (SAScore) suggest these compounds are amenable to chemical synthesis, facilitating experimental validation and iterative medicinal chemistry optimization.

4.1 Limitations

The primary limitation of this study is the exclusive reliance on computational methods. While the integrated pipeline provides a powerful approach for hypothesis generation and candidate prioritization, the predictions require experimental validation. The identified candidate compounds have not been tested in biochemical binding assays, cellular functional assays or DDLPS-relevant cell lines or animal models. The predicted IC₅₀ values (<10 nM), docking scores, and MM/GBSA binding free energies are computational estimates that may substantially differ from experimentally measured values. MM/GBSA calculations are known to overestimate binding affinities for lipophilic ligands, and the absolute energies reported should not be interpreted as quantitative predictions of experimental *K_d* or IC₅₀ values. The docking scores, while useful for comparative ranking, are approximations that may not fully capture the thermodynamics of binding. The network pharmacology analysis is inherently correlative. The predicted targets, while enriched

in p53 and cell-cycle pathways, have not been confirmed through target engagement assays or cellular thermal shift assays. The protein-protein interaction network is based on database-derived interactions that may not reflect the specific biology of DDLPS, and the functional consequences of polypharmacological engagement remain speculative. The transcriptomic data used for network pharmacology (GSE21122) provides a valuable but static snapshot of gene expression and may not capture the full complexity of DDLPS biology, including intratumoral heterogeneity and the influence of the tumor microenvironment. The ADMET predictions, while indicating favorable properties, are based on machine-learning models trained on existing datasets and may not accurately predict the behavior of structurally novel compounds. The toxicity alerts identified require experimental confirmation, and the actual safety profile may differ substantially from predictions. The molecular dynamics simulations were performed for 500 ns, which may be insufficient to fully sample conformational space for these protein-ligand complexes. While the RMSD and interaction analysis suggested stable binding, longer simulations would provide greater confidence in the binding mode predictions. Finally, the study focused on the p53-binding domains of MDM2 and MDM4 and did not consider other functional domains or the potential for allosteric modulation. While the p53-binding cleft is the primary target for small-molecule inhibitors, other mechanisms of MDM2/MDM4 inhibition may also be relevant.

4.2 Future directions

Future work could focus on experimental validation of binding affinities for ZINC9412756 and ZINC20724474 against recombinant MDM2 and MDM4 proteins using surface plasmon resonance, fluorescence polarization, or isothermal titration calorimetry. Should these studies confirm binding, subsequent evaluation of p53 pathway activation and antiproliferative activity in DDLPS cell lines that model the MDM2-amplified, wild-type *TP53* genotype would be a logical next step, including assessment of p53 protein levels, transcriptional targets, and cell-cycle/apoptosis markers. Assessment of selectivity against a panel of related and unrelated targets would also be valuable to evaluate target specificity. If the candidate compounds demonstrate promising activity in initial assays, iterative structure-activity relationship optimization guided by the docking and MD interaction models could be pursued to improve potency, selectivity, and ADMET properties while mitigating the toxicity alerts identified in this study. Experimental ADMET profiling would be necessary to confirm or refine the in silico predictions, including metabolic stability, plasma protein binding, CYP inhibition, and safety pharmacology studies. Should the candidates demonstrate sufficient in vitro activity and favorable pharmacokinetic properties, evaluation in DDLPS xenograft models could be considered as a longer-term objective.

5.0 Conclusion

In this study, we employed an integrated computational pipeline to prioritize candidate dual MDM2/MDM4 inhibitors for DDLPS. Virtual screening of 9.4 million drug-like compounds identified ZINC9412756 and ZINC20724474 as candidate compounds with predicted affinity for both MDM2 and MDM4, driven predominantly by hydrophobic interactions. Structural similarity analysis against known MDM2 (n = 4,154) and MDM4 (n = 155) inhibitors indicated that both candidates are structurally distinct from previously reported chemotypes, with maximum ECFP4 Tanimoto similarities of 0.3827 and 0.3488 to the closest MDM2 inhibitors and 0.2500 and 0.2315 to the closest MDM4 inhibitors. These findings suggest that ZINC9412756 and ZINC20724474 represent novel chemotypes for dual MDM2/MDM4 inhibition that warrant further investigation. Network pharmacology predicted engagement of a DDLPS-relevant module enriched in p53 signaling, providing a mechanistic rationale for further investigation. ADMET profiling predicted several toxicity alerts (hepatotoxicity, nephrotoxicity, hERG liability) that represent tractable optimization targets, with ZINC9412756 emerging as the more balanced candidate based on its lower hERG liability and P-gp substrate potential. The key limitations of this study include the likely overestimation of binding affinities by MM/GBSA, the correlative nature of the network pharmacology analysis, and the absence of experimental validation. The candidate compounds require biochemical binding assays, cellular activity studies in DDLPS models, and experimental ADMET profiling to confirm their potential. This integrated computational approach has identified promising candidate compounds for further experimental evaluation as dual MDM2/MDM4 inhibitors. The pipeline suggests that combining multiple computational modalities can efficiently navigate vast chemical spaces, offering a broadly applicable strategy for prioritizing candidates against complex oncogenic targets. Future work will focus on experimental validation of the top candidate compounds and iterative structure-activity relationship optimization to address the identified toxicity alerts.

6.0 Credit Authorship Contribution Statement

Prince Danan Biniyam: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization.

Solomon Sackitey: Methodology, Software, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization.

Alberta Hammah: Methodology, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing.

Charles Lamptey Mireku: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing

Jesse Ayim Buabeng: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing.

Josephine Frimpomaa Oppong: Methodology, Validation, Writing - Review & Editing.

Lawrence Kwame Nkrumah: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing

Kwabena Adu-Adjei: Methodology, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing.

Wise Mawuli Asare: Methodology, Validation, Writing - Review & Editing.

Clifford Penajah: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing.

Emmanuel Gyamfi Obeng: Methodology, Validation, Writing - Review & Editing

Cedric Dzidzor Kodjo Amengor: Supervision, Project Administration, Resources, Writing - Review & Editing.

7.0 Acknowledgements

The authors would like to express their sincere gratitude to the computational drug discovery unit at the School of Pharmacy, University of Health and Allied Sciences, for their invaluable support.

8.0 Data and Software Availability

All data generated and analyzed during this study are available in the main manuscript and the Supporting Information. The Python scripts used for QSAR model development are available at [<https://github.com/Biniyam1997/MDM2>].

9.0 Ethics declaration

Not applicable.

10.0 Funding

This work received no funding.

References

1. Gamboa, A. C., Gronchi, A., & Cardona, K. (2023). Soft-tissue sarcoma in adults: An update on histiotype-specific management in precision medicine era. *CA: A Cancer Journal for Clinicians*, 73(2), 189-215. <https://doi.org/10.3322/caac.21778>
2. Crago, A. M., & Dickson, M. A. (2016). Liposarcoma: Multimodality management and future targeted therapies. *Surgical Oncology Clinics of North America*, 25(4), 761-773.
3. Singer, S., Antonescu, C. R., Riedel, E., & Brennan, M. F. (2003). Histologic subtype and margin of resection predict pattern of recurrence and survival for retroperitoneal liposarcoma. *Annals of Surgery*, 238(3), 358-371.
4. Mussi, C., Collini, P., Miceli, R., Barisella, M., Mariani, L., Fiore, M., Casali, P. G., & Gronchi, A. (2008). The prognostic impact of dedifferentiation in retroperitoneal liposarcoma: A series of surgically treated patients at a single institution. *Cancer*, 113(7), 1657-1665. <https://doi.org/10.1002/cncr.23774>
5. Italiano, A., Toulmonde, M., Cioffi, A., Penel, N., Isambert, N., Bompas, E., Duffaud, F., Patrikidou, A., Lortal, B., Le Cesne, A., Blay, J. Y., Maki, R. G., Schwartz, G. K., Antonescu, C. R., Singer, S., Coindre, J. M., & Bui, B. (2012). Advanced well-differentiated/dedifferentiated liposarcomas: Role of chemotherapy and survival. *Annals of Oncology*, 23(6), 1601-1607. <https://doi.org/10.1093/annonc/mdr428>
6. Livingston, J. A., Bugano, D., Barbo, A., et al. (2017). Role of chemotherapy in dedifferentiated liposarcoma of the retroperitoneum: Defining the benefit and challenges of the standard. *Scientific Reports*, 7, Article 11836. <https://doi.org/10.1038/s41598-017-12132-6>
7. Gamboa, A. C., Gronchi, A., & Cardona, K. (2020). Soft-tissue sarcoma in adults: An update on the current state of histiotype-specific management in an era of personalized medicine. *CA: A Cancer Journal for Clinicians*, 70(3), 200-229. <https://doi.org/10.3322/caac.21605>
8. Barretina, J., Taylor, B. S., Banerji, S., et al. (2010). Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nature Genetics*, 42(8), 715-721. <https://doi.org/10.1038/ng.619>
9. Leach, F. S., Tokino, T., Meltzer, P., et al. (1993). p53 mutation and MDM2 amplification in human soft tissue sarcomas. *Cancer Research*, 53(10), 2231-2234.
10. Espadinha, M., Santos, F., Neves, J. D. L., & Castanheiro, M. (2022). Discovery of MDM2-p53 and MDM4-p53 protein-protein interactions small molecule dual inhibitors. *European Journal of Medicinal Chemistry*, 241, 114637. <https://doi.org/10.1016/j.ejmech.2022.114637>
11. Li, K., Zhang, Y., Wang, L., & Chen, J. (2024). Searching for novel MDM2/MDMX dual inhibitors through a drug repurposing approach. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 39(1), 2288810. <https://doi.org/10.1080/14756366.2023.2288810>

12. Toledo, F., & Wahl, G. M. (2023). Regulating the p53 pathway: Therapeutic opportunities and challenges. *Nature Reviews Cancer*, 23(8), 512-528. <https://doi.org/10.1038/s41568-2023-00589-w>
13. Jones, S. N., Roe, A. E., Donehower, L. A., & Bradley, A. (1995). Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53. *Nature*, 378(6553), 206-208. <https://doi.org/10.1038/378206a0>
14. Montes de Oca Luna, R., Wagner, D. S., & Lozano, G. (1995). Rescue of early embryonic lethality in mdm2-deficient mice by deletion of p53. *Nature*, 378(6553), 203-206. <https://doi.org/10.1038/378203a0>
15. Oliner, J. D., Kinzler, K. W., Meltzer, P. S., George, D. L., & Vogelstein, B. (1992). Amplification of a gene encoding a p53-associated protein in human sarcomas. *Nature*, 358(6381), 80-83. <https://doi.org/10.1038/358080a0>
16. Vassilev, L. T., Vu, B. T., Graves, B., et al. (2004). In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*, 303(5659), 844-848. <https://doi.org/10.1126/science.1092472>
17. Tovar, C., Rosinski, J., Filipovic, Z., et al. (2006). Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: Implications for therapy. *Proceedings of the National Academy of Sciences*, 103(6), 1888-1893. <https://doi.org/10.1073/pnas.0507493103>
18. Vu, B., Wovkulich, P., Pizzolato, G., et al. (2013). Discovery of RG7112: A small-molecule MDM2 inhibitor in clinical development. *ACS Medicinal Chemistry Letters*, 4(5), 466-469. <https://doi.org/10.1021/ml4000657>
19. Konopleva, M., Martinelli, G., Daver, N., et al. (2020). MDM2 inhibition: An important step forward in cancer therapy. *Leukemia*, 34(11), 2858-2874. <https://doi.org/10.1038/s41375-020-0949-z>
20. Wang, S., Sun, W., Zhao, Y., et al. (2014). SAR405838: An optimized inhibitor of MDM2-p53 interaction that induces complete and durable tumor regression in cancer models. *Cancer Research*, 74(20), 5855-5865. <https://doi.org/10.1158/0008-5472.CAN-14-0799>
21. Ding, Q., Zhang, Z., Liu, J. J., et al. (2013). Discovery of RG7388, a potent and selective p53-MDM2 inhibitor in clinical development. *Journal of Medicinal Chemistry*, 56(14), 5979-5983.
22. Sun, D., Li, Z., Rew, Y., et al. (2014). Discovery of AMG 232, a potent, selective, and orally bioavailable MDM2-p53 inhibitor in clinical development. *Journal of Medicinal Chemistry*, 57(4), 1454-1472.
23. Stein, E. M., DeAngelo, D. J., Chromik, J., et al. (2022). Results from a first-in-human phase I study of siremadlin (HDM201) in patients with advanced wild-type TP53 solid tumors and acute leukemia. *Clinical Cancer Research*, 28(5), 870-881.
24. Ray-Coquard, I., Blay, J. Y., Italiano, A., et al. (2012). Effect of the MDM2 antagonist RG7112 on the p53 pathway in patients with MDM2-amplified, well-differentiated or dedifferentiated liposarcoma: An exploratory proof-of-mechanism study. *The Lancet Oncology*, 13(11), 1133-1140.

25. de Jonge, M., de Weger, V. A., Dickson, M. A., et al. (2017). A phase I study of SAR405838, a novel human double minute 2 (HDM2) antagonist, in patients with solid tumours. *European Journal of Cancer*, 76, 144-151.
26. Frontiers in Oncology. (2022). Targeting the MDM2-p53 pathway in dedifferentiated liposarcoma. *Frontiers in Oncology*, 12, 1006959. <https://doi.org/10.3389/fonc.2022.1006959>.
27. Aziz, M. H., Shen, H., & Maki, C. G. (2011). Acquisition of p53 mutations in response to the non-genotoxic p53 activator Nutlin-3. *Oncogene*, 30(46), 4678-4686.
28. Wade, M., Wang, Y. V., & Wahl, G. M. (2010). The p53 orchestra: Mdm2 and Mdmx set the tone. *Trends in Cell Biology*, 20(5), 299-309.
29. Patton, J. T., Mayo, L. D., Singhi, A. D., Gudkov, A. V., Stark, G. R., & Jackson, M. W. (2006). Levels of HdmX expression dictate the sensitivity of normal and transformed cells to Nutlin-3. *Cancer Research*, 66(6), 3169-3176.
30. Toledo, F., & Wahl, G. M. (2006). Regulating the p53 pathway: In vitro hypotheses, in vivo veritas. *Nature Reviews Cancer*, 6(12), 909-923.
31. Giachetti, C., Michelacci, S., & Zannini, M. (2023). MDM4/MDMX as critical regulators of p53 in cancer: New therapeutic avenues. *Trends in Cell Biology*, 33(6), 478-492. <https://doi.org/10.1016/j.tcb.2023.02.008>
32. Li, K., Zhang, Y., Wang, L., & Chen, J. (2024). Searching for novel MDM2/MDMX dual inhibitors through a drug repurposing approach. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 39(1), 2288810. <https://doi.org/10.1080/14756366.2023.2288810>
33. Giachetti, C., Michelacci, S., & Zannini, M. (2023). MDM4/MDMX as critical regulators of p53 in cancer: New therapeutic avenues. *Trends in Cell Biology*, 33(6), 478-492. <https://doi.org/10.1016/j.tcb.2023.02.008>
34. Li, K., Zhang, Y., Wang, L., & Chen, J. (2024). Searching for novel MDM2/MDMX dual inhibitors through a drug repurposing approach. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 39(1), 2288810. <https://doi.org/10.1080/14756366.2023.2288810>
35. Pellegrino, M., Mancini, F., Lucà, R., et al. (2015). Targeting the MDM2/MDM4 interaction interface as a promising approach for p53 reactivation therapy. *Cancer Research*, 75(21), 4560-4572.
36. Zhang, S., Yan, Z., Li, Y., Gong, Y., Lyu, X., Lou, J., Zhang, D., Meng, X., & Zhao, Y. (2022). Structure-based discovery of MDM2/4 dual inhibitors that exert antitumor activities against MDM4-overexpressing cancer cells. *Journal of Medicinal Chemistry*, 65(8), 6207-6230.
37. Vachani, P., Mascarenhas, J., et al. (2021). A phase 1b/2 study of navtemadlin (KRT-232), a murine double minute 2 inhibitor, combined with a BCR-ABL tyrosine kinase inhibitor in patients with relapsed/refractory, TP53 wild-type, Ph+ chronic myeloid leukemia. *Blood*, 138(Suppl 1), Article 4983.

38. Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., & Sayle, R. A. (2020). ZINC20—A free ultralarge-scale chemical database for ligand discovery. *Journal of Chemical Information and Modeling*, 60(12), 6065-6073.
39. Salama, M. M., Shafaa, M. W., El-Nagdy, M. E., & Hasan, M. E. (2025). Bioinformatics approach for discovery of potential lead compound of NSP6 of SARS-CoV-2 using structure based virtual screening and molecular dynamics simulations. *Scientific Reports*, 15, Article 37347.
40. Fouad, M. A., Osman, A. A., Abdelhamid, N. M., Rashad, M. W., Nabawy, A. Y., & El Kerdawy, A. M. (2024). Discovery of dual kinase inhibitors targeting VEGFR2 and FAK: Structure-based pharmacophore modeling, virtual screening, and molecular docking studies. *BMC Chemistry*, 18, Article 30.
41. Cherkasov, A., Muratov, E. N., Fourches, D., et al. (2014). QSAR modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry*, 57(12), 4977-5010.
42. Mendez, D., Gaulton, A., Bento, A. P., et al. (2019). ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Research*, 47(D1), D930-D940.
43. Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., & Moreno, E. (2021). gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS. *Journal of Chemical Theory and Computation*, 17(10), 6281-6291.
44. Landrum, G. (2016). *RDKit: Open-source cheminformatics software*. <http://www.rdkit.org>
45. Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, 29(6-7), 476-488.
46. Rogers, D., & Hahn, M. (2010). Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5), 742-754.
47. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
48. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
49. Ke, G., Meng, Q., Finley, T., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146-3154.
50. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
51. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42.
52. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232.
53. Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455-461.

54. Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. *Methods in Molecular Biology*, 1263, 243-250.
55. Pettersen, E. F., Goddard, T. D., Huang, C. C., et al. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605-1612.
56. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, 27(3), 221-234.
57. Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11), 1859-1865.
58. Huang, J., & MacKerell, A. D., Jr. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, 34(25), 2135-2145.
59. Abraham, M. J., Murtola, T., Schulz, R., et al. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, 19-25.
60. Szklarczyk, D., Kirsch, R., Koutrouli, M., et al. (2023). The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638-D646.
61. Daina, A., Michielin, O., & Zoete, V. (2019). SwissTargetPrediction: Updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Research*, 47(W1), W357-W364.
62. Stelzer, G., Rosen, N., Plaschkes, I., et al. (2016). The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics*, 54, 1.30.1-1.30.33.
63. Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, 7, Article 42717.
64. Fu, L., Shi, S., Yi, J., et al. (2024). ADMETlab 3.0: An updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Research*, 52(W1), W422-W431.
65. Biniyam, P. D., Kayeri, N., Buabeng, J. A., et al., & Amengor, C. D. K. (2026). Discovery of novel MDM2 inhibitors from a *Penicillium* metabolome library: An integrated phylogenetic, machine learning, and molecular simulation approach. *Journal of Molecular Graphics and Modelling*. Advance online publication.
66. Biniyam, P. D., Aboagye, D., Nkrumah, L. K., Sackitey, S., Buabeng, J. A., Ohene-Adu, V., ... & Amengor, C. D. K. (2026). Dual-Targeting MDM2 and CDK4: An Integrated Machine Learning and Molecular Simulation Pipeline Identifies Novel Fungal Metabolites for Cancers with Co-Dysregulated p53 and Rb Pathways. *In Silico Research in Biomedicine*, 100340.

67. Biniyam, P. D., Kayeri, N., Ayihi, S. J., Haider, S. A., Dzorwoko, H., Adu-Adjei, K., ... & Amengor, C. D. K. (2026). A Computational Pipeline Combining Machine Learning and Molecular Simulations Identifies Repurposed PI3K α Inhibitors from FDA-Approved Drugs. *Artificial Intelligence Chemistry*, 100111.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInfo.docx](#)